Cognitive Psychology xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

A pessimistic view of optimistic belief updating

Punit Shah [a,b], Adam J.L. Harris [c,*], Geoffrey Bird [b,d], Caroline Catmur [e,f], Ulrike Hahn [a]

[a] Department of Psychological Sciences, Birkbeck College, University of London, Malet Street, London WC1E 7HX, United Kingdom
[b] MRC Social, Genetic, & Developmental Psychiatry Centre, De Crespigny Park, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AF, United Kingdom
[c] Department of Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, United Kingdom
[d] Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, United Kingdom
[e] Department of Psychology, University of Surrey, Guildford GU2 7XH, United Kingdom
[f] Department of Psychology, De Crespigny Park, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AF, United Kingdom

ARTICLE INFO

Article history:
Accepted 25 May 2016
Available online xxxx

Keywords:
Unrealistic optimism
Optimism bias
Motivated reasoning
Human rationality
Belief updating
Bayesian belief updating

ABSTRACT

Received academic wisdom holds that human judgment is characterized by unrealistic optimism, the tendency to underestimate the likelihood of negative events and overestimate the likelihood of positive events. With recent questions being raised over the degree to which the majority of this research genuinely demonstrates optimism, attention to possible mechanisms generating such a bias becomes ever more important. New studies have now claimed that unrealistic optimism emerges as a result of biased belief updating with distinctive neural correlates in the brain. On a behavioral level, these studies suggest that, for negative events, desirable information is incorporated into personal risk estimates to a greater degree than undesirable information (resulting in a more optimistic outlook). However, using task analyses, simulations, and experiments we demonstrate that this pattern of results is a statistical artifact. In contrast with previous work, we examined participants' use of new information with reference to the normative, Bayesian standard. Simulations reveal the fundamental difficulties that would need to be overcome by any robust test of optimistic updating. No such test presently exists, so that the best one can presently do is perform analyses with a number of techniques, all of which have important weaknesses. Applying these analyses to five experiments shows no evidence of optimistic updating. These results clarify the difficulties involved in studying

* Corresponding author.
E-mail address: adam.harris@ucl.ac.uk (A.J.L. Harris).

http://dx.doi.org/10.1016/j.cogpsych.2016.05.004
0010-0285/

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

human 'bias' and cast additional doubt over the status of optimism as a fundamental characteristic of healthy cognition.

1. Introduction

For over 30 years it has been an accepted 'fact' that humans are subject to a consistent bias when estimating personal risk. Research suggests that people underestimate their chances of experiencing negative events (with respect to their estimates of the average person's risk), and overestimate their chances of experiencing positive events (e.g., Harris & Guten, 1979; Weinstein, 1980, 1982, 1984, 1987). Hence researchers in this area have concluded that "people have an optimistic bias concerning personal risk" (Weinstein, 1989, p. 1232). This pattern of optimistic self-estimates has been termed 'unrealistic optimism', and is commonly thought to reflect a self-serving motivational bias (for a review, see Helweg-Larsen & Shepperd, 2001; but see also Chambers & Windschitl, 2004).

Unrealistic optimism has attracted a great deal of academic interest both from multiple domains within psychology (including social psychology, judgment and decision-making, and cognitive neuroscience) and from economics (see e.g., van den Steen, 2004). This research has also been used in various applied domains including clinical psychology where it has been proposed that an optimistic bias is a necessary requirement to guard against depression (see Taylor & Brown, 1988). Within health psychology, unrealistic optimism is used to explain the failure of individuals to undertake health protective behaviors (e.g., van der Velde, Hooykaas, & van der Joop, 1992; van der Velde, van der Pligt, & Hooykaas, 1994) and to resist changes in diet (Shepherd, 2002), on the grounds that personal risk estimates of obesity-related diseases are underestimated (see Miles & Scaife, 2003). Within the financial sector, unrealistic optimism has been linked to economic choice (HM Treasury Green Book, n.d.; Puri & Robinson, 2007; Sunstein, 2000) and it has been suggested as one of the factors behind the financial crisis experienced in the first decade of the 21st Century (Sharot, 2012). Most recently, attention has turned to investigating the neural correlates underlying the phenomenon (Chowdhury, Sharot, Wolfe, Düzel, & Dolan, 2014; Garrett et al., 2014; Sharot, Guitart-Masip, Korn, Chowdhury, & Dolan, 2012; Sharot, Kanai, et al., 2012; Sharot, Korn, & Dolan, 2011; Sharot, Riccardi, Raio, & Phelps, 2007; see Sharot, 2012).

1.1. Detecting optimism: the comparison method

A recent analysis, however, has cast doubt over the evidential basis for unrealistic optimism. Harris and Hahn (2011) argued that methodological and conceptual limitations of studies investigating this phenomenon mean that results may be better explained as a statistical artifact rather than unrealistic optimism (for a critique and counter-critique, see Hahn & Harris, 2014; Shepperd, Klein, Waters, & Weinstein, 2013). Harris and Hahn demonstrated that it was possible for perfectly rational (non-optimistic) agents to generate personal risk estimates that would be classified as unrealistically optimistic given the paradigms and scoring methods used in the vast majority of unrealistic optimism studies. Specifically, unrealistic optimism is usually studied by asking participants to compare (directly or indirectly) their chance of experiencing a negative life event with the chance of the average individual ('*the comparison method*'). The typical result is that, at a group level, participants' average estimates of their own risk are significantly lower than the group average. Harris and Hahn, however, showed that when the negative events are rare (i.e., have a base rate of less than 50% within the population, as is almost always the case in optimism studies), three statistical factors, namely, attenuated response scales, under-sampling of population minorities, and regressive population base rate estimates, can cause completely rational groups of agents to produce the pattern of empirical results that has been taken to indicate unrealistic optimism. This methodological failing means that the results of past studies using the comparison method (i.e., the majority of research on optimism

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

to date) cannot be taken as genuine evidence of an optimistic bias. Whether or not people *are* optimistically biased can no longer be considered a settled question. Harris and Hahn (2011) thus suggest that rather than being a distinguishing feature of healthy human thought (e.g., Sharot, 2012; Taylor & Brown, 1988), 'unrealistic optimism' may purely be a statistical artifact resulting from flawed empirical methodologies.

Crucially, the statistical artifact account is valence *in*dependent; it relies solely on the frequency of the events to be judged and not on whether the effect equates to optimism or pessimism. Therefore judgments in which one's own chance is estimated to be lower than the average person's chance should also be observed when relatively rare *positive* events are estimated. However, because experiencing positive events is a desirable result, the same pattern of responding would traditionally be interpreted as unrealistic *pessimism*. In contrast, an unrealistic optimism account would suggest that one's own 'risk' of experiencing positive events is judged to be higher than that of the average person. Thus, the inclusion of rare positive events is a critical test for distinguishing genuinely optimistic responding from potentially artifactual optimism using the comparison method. Studies that have included such events have found a pattern of responding that is inconsistent with an optimistic bias, but is consistent with the statistical artifact account: lower estimates of one's own risk for *both* positive and negative events (Chambers, Windschitl, & Suls, 2003; Harris, 2009; Kruger & Burrus, 2004; Moore & Small, 2008).

1.2. The update method

The majority of evidence for unrealistic optimism is based on the flawed comparison method. This means that, despite the considerable amount of research on the topic over the last 30 years, further empirical work is required to firmly establish the phenomenon. Moreover, even if unrealistic optimism does exist, the use of the flawed comparison method in the majority of research aimed at understanding the factors that influence it means that we have considerably less knowledge about its potential causes and moderators than widely thought.

It is therefore of note that a recent series of high-profile studies (Chowdhury et al., 2014; Garrett & Sharot, 2014; Garrett et al., 2014; Korn, Sharot, Walter, Heekeren, & Dolan, 2014; Kuzmanovic, Jefferson, & Vogeley, 2015, 2016; Moutsiana et al., 2013; Sharot, Guitart-Masip, et al., 2012; Sharot, Kanai, et al., 2012; Sharot et al., 2011) has purported to extend the understanding of unrealistic optimism by investigating how an optimistic bias might be maintained. In these studies, pioneered by Sharot and colleagues, researchers asked their participants to estimate their chance of experiencing a series of negative events and then gave them the population base rates (hereafter 'base rates') of those events, that is, participants were told the probability with which these negative events are experienced by the average individual. Subsequently, participants were asked to re-estimate their own chance of experiencing the negative life events. The degree to which participants updated their personal risk estimates (i.e., the difference between their initial and second estimate of personal risk) was measured. Participants updated their estimates significantly more in response to desirable information (information suggesting that the base rate of the negative event, and hence average person's risk, was lower than the participant's personal risk estimate) than they updated their estimates in response to undesirable information (information suggesting that the base rate of the negative event was higher than that estimated by the participant). This pattern of results led the researchers to infer that participants were selectively incorporating new information in order to maintain an optimistic outlook. Functional Magnetic Resonance Imaging (fMRI) revealed that activity in right inferior frontal gyrus predicted updating in response to undesirable information, while activity in medial frontal cortex/superior frontal gyrus and right cerebellum predicted updating in response to desirable information (Sharot et al., 2011).

1.3. Existence of unrealistic optimism

These recent findings on belief updating are of great interest for two reasons. Firstly, in the light of critiques of unrealistic optimism research (Harris & Hahn, 2011), these results might be seen to tip the balance of evidence further toward the widespread presence of unrealistic optimism in everyday life. If people revise their beliefs more in response to desirable than undesirable information, this would nec-

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

essarily give rise to unrealistic optimism to the extent that it would make people consider positive events more likely to happen to them than negative events of equal probability. Information that lowers the probability of a negative event is desirable, but that same lowering is undesirable in the context of positive events (which we *do* want to experience). Selective updating in response to desirable as opposed to undesirable information would thus necessarily leave estimates of otherwise matched positive events higher than their negative counterparts. This in itself would constitute unrealistic optimism (Lench & Ditto, 2008). In other words, selectively underweighting undesirable information relative to desirable information will lower estimates of negative events and inflate estimates of positive events. This would also make it surprising if unrealistic optimism were not observed in future comparative tests of optimism that effectively control for the confounds identified in Harris and Hahn (2011).

1.4. Mechanisms of unrealistic optimism

Secondly, Sharot et al. provide evidence for a mechanism by which unrealistic optimism might emerge or persist. It has long been held that unrealistic optimism reflects motivated reasoning that serves to promote psychological (Taylor & Brown, 1988, 1994) and physical (Sharot, 2012) well-being. Yet it has often been left unspecified what form exactly such motivated reasoning might take. Hence the question of mechanism has continued to loom large. Some researchers have argued for non-motivated mechanisms leading to unrealistic optimism, such as egocentricism (e.g., Chambers et al., 2003) and 'differential regression' (Moore & Small, 2008; see also, Moore & Healy, 2008). Though these mechanisms may on occasion give rise to optimism, because they do not reflect a motivational, or valence-based, bias, they may in other cases give rise to pessimism. This challenges the contention that people are generally over-optimistic. It also challenges the contention that optimism promotes well-being, as both optimism and pessimism are then secondary characteristics of healthy human thought. There have been some more detailed accounts of motivated reasoning (e.g., Critcher & Dunning, 2009; Dunning, Meyerowitz, & Holzberg, 1989; Lench & Bench, 2012), though much of their application has been in other domains, such as self-perceptions of skill or attractiveness, not optimism about future life events (but see e.g., Ditto, Jemmott, & Darley, 1988; Ditto & Lopez, 1992; Lench & Ditto, 2008). Moreover, many of these accounts view the impact of motivation in guiding the depth and scope of cognitive processing, not as a directly biasing force per se (see also Kunda, 1990). This restricts the circumstances in which optimistic conclusions will be attainable, and thus sits uneasily both with the idea that unrealistic optimism is a pervasive bias and that it exists because it has adaptive value (see also, Hahn & Harris, 2014). As a consequence, selective belief updating potentially provides a long-missing, fully specified, process account of how unrealistic optimism might come to be, in addition to providing support for the very existence of the phenomenon itself. Thus, for both skeptics and champions of unrealistic optimism, the degree to which people show evidence of optimistic belief updating is of considerable interest.

1.5. Overview of the present paper

The current paper provides a detailed analysis of the claim for optimistic belief updating in the form of: computational task analyses, simulations, and five experiments. The task analyses and simulations highlight flaws in the rationale of the update methodology. The experimental results demonstrate that these flaws are consequential in experiments with human participants. The results from these lines of enquiry converge to demonstrate that reports of optimistic updating reflect statistical artifacts, rather than genuinely optimistic belief updating (as has been argued for demonstrations of unrealistic optimism using the comparison method, Harris & Hahn, 2011).

Experiment 1 (Section 2) modifies the update method, first introduced in Sharot et al. (2011), to examine updating in response to desirable and undesirable information for both negative *and* positive events. Observed patterns of updating conform to the predictions of the 'statistical artifact account', *not* optimistic updating. The possibility of an artifactual explanation for seemingly optimistic belief updating is then pursued further by considering the way in which rational agents *should* update their beliefs in response to new information, and then exploring the consequences of this for the update method. It is demonstrated that the version of the update method used by Sharot and colleagues is

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

normatively inappropriate (Sections 2.3.1 and 2.3.2). Simulations demonstrate how a pattern of 'biased' belief updating can be obtained from a population of rational agents (Section 3). These simulations also highlight the difficulty of conducting *any* robust test of bias in belief updating concerning likelihood estimates for future life events. In light of these difficulties, the best one can presently do is perform analyses with a number of techniques, all of which have important weaknesses. Experiments 2, 3 (A & B) and 4 conduct such analyses (Sections 4–7), yielding no support for the notion of optimistically biased belief updating.

2. Experiment 1

Experiment 1 provided a partial replication of an experiment using the update method (e.g., Sharot et al., 2011). The update method was used to test for unrealistic optimism, but with two primary modifications. The first modification was the addition of positive events. Harris and Hahn (2011) showed how the inclusion of positive events provides a simple test for artifactual optimism in the context of the comparison method. For the new update method, the addition of positive events provides a similar test for potential artifacts. Information is desirable with respect to negative events when it suggests events are *less* probable than previously estimated. The converse is true for positive events; information is desirable when it suggests events are *more* probable than previously estimated. Changing one's beliefs less in response to information suggesting the event is less probable than previously estimated, and consequently under-estimating one's personal chance of experiencing the event in question, thus signals optimism only for negative events; for positive events, such under-estimates signal pessimism. If belief updating truly reflects unrealistic optimism then it should be greater in response to desirable information for both positive and negative events, even though this involves deviation from the initial self estimate in opposite directions.

Readers familiar with Sharot et al. (2011; see also, Garrett & Sharot, 2014; Garrett et al., 2014; Korn et al., 2014; Moutsiana et al., 2013; Sharot, Kanai, et al., 2012) might point out that positive events were included. Sharot and colleagues created positive events through asking participants to judge the complementary likelihood of *not* experiencing each negative event, and found similar optimistic updating to that found with the original wording. Such a result does not provide the critical test required, however, since these 'positive events' will be exact complements of the negative events. Consequently, any statistical mechanism that might be exerting an influence on the results of the negative events should exert exactly the opposite influence on *these* 'positive events' (e.g., if the overestimation of rare events is a contributing factor, then the complementary probability will be underestimated) and the same pattern of results is predicted to be observed on any theory. This point will be further clarified in discussion of the simulation data that follow (Section 3). A collection of different, genuinely positive, events is thus required.

The second modification we introduced concerned the base rates that were presented to participants. Sharot and colleagues (e.g., Sharot et al., 2011) used probabilities obtained from sources such as the UK's Office for National Statistics. These sources tend to be focused almost exclusively on negative events such as disease and divorce, making it difficult to obtain statistics for positive events. Furthermore, we wished to *manipulate* the desirability of information presented to participants for maximum experimental control. The base rates presented to participants in Experiment 1 were therefore derived from participants' initial self-estimates (hereafter SE1s) for both positive and negative events (see Section 2.1.3.1; see also, Kuzmanovic et al., 2015, 2016). A funneled debrief procedure (Bargh & Chartrand, 2000) was used to ensure that any participants who suspected that the probabilities might be inaccurate were removed from the analysis. For consistency with previous literature, and with Experiment 3 (Sections 5 and 6) which used externally sourced probabilities, the derived probabilities of Experiments 1 and 2 will be referred to as "actual" probabilities.

2.1. Method

2.1.1. Participants

Thirteen healthy participants (6 females; aged 19–28 [median = 20]) were recruited via the Birkbeck Psychology participant database. Two additional participants were recruited and tested but were

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

not included in the analysis as the funneled debrief procedure revealed that they were surprised by some of the base rates that were presented to them.[1] Due to the association between unrealistic optimism and depression (e.g., Strunk, Lopez, & DeRubeis, 2006), all participants were screened for depression using the Beck Depression Inventory-II (Beck, Steer, & Brown, 1996) before completing the study. On this measure no participant met accepted criteria for depression (M = 3.2, SE = 0.7). All participants gave informed consent and were paid for their participation.

2.1.2. Stimuli

Eighty short descriptions of life events (see Appendix A), many of which had previously been used in the study of unrealistic optimism (Lench & Ditto, 2008; Sharot, Guitart-Masip, et al., 2012; Sharot, Kanai, et al., 2012; Sharot et al., 2011; Weinstein, 1980), were presented in a random order. Half of the events were positive and half negative. We limited the number of very rare or very common events. The events for which base rates were available lay between 10% and 70% (M = 32.6, SD = 18.8; Office for National Statistics and PubMed), providing participants with the opportunity to underestimate and overestimate the likelihood of each event.

2.1.3. Procedure

The trial structure is shown in Fig. 1A. Each trial began with the presentation of text describing a life event for 5 s, during which time participants were asked to imagine that event. Participants were then instructed to estimate the likelihood of that event happening to them using a computer keyboard (the Initial Self Estimate; SE1). If the participant did not respond within 5 s, the trial was omitted from analysis (M = 1.9 trials per participant, SD = 2.0). A fixation cross was then displayed for 1 s, followed by presentation of the event description accompanied by the base rate. This was described to participants as the likelihood of that event occurring at least once to a person living in the same socio-cultural environment as the participant (on derivation of base rates see 'Section 2.1.3.1'). A second fixation cross appeared for 1 s, after which the next trial began. Participants were instructed that if they saw events which they had already experienced in their lifetime, they should estimate the likelihood of that event happening to them again in the future. Eighty trials were presented in a random order, comprising 20 trials involving positive life events accompanied by desirable information concerning their likelihood, 20 trials involving positive life events accompanied by undesirable information concerning their likelihood, 20 trials involving negative life events accompanied by desirable information concerning their likelihood, and 20 trials involving negative life events accompanied by undesirable information concerning their likelihood. These trials were preceded by two practice trials.

Participants estimated the probability of each event twice (see Appendix B, Table B1, for mean estimates at each stage): once during the session described above and once in a second session (henceforth, Second Estimate; SE2) that immediately followed the first, in which they were again asked to estimate the probability of the event. The difference between SE1 and SE2 was taken as the measure of the degree to which participants had updated their judgment of the event occurring to them in their lifetime. The particular events associated with desirable or undesirable information were randomized across participants.

2.1.3.1. Base rates. For each event, the average probability of that event occurring at least once to a person living in the same socio-cultural environment as the participant was derived from the participant's SE1 and presented to the participant as the actual event base rate. Probabilities were computed according to the following formula: A random percentage between 17% and 40% (uniform distribution) of the SE1 was either added to, or subtracted from, the SE1, according to trial type, and rounded to the nearest integer. Thus, on positive desirable trials a random percentage of the SE1 was added to the SE1, resulting in a derived probability indicating that the positive event was more likely to occur than had previously been estimated. On positive undesirable trials a random percentage of the SE1 was subtracted from the SE1, indicating that the positive event was less likely than had been estimated. On negative desirable trials a random percentage of the SE1 was subtracted from the SE1,

[1] Including these participants in the analysis does not change the pattern, or statistical significance, of the results.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

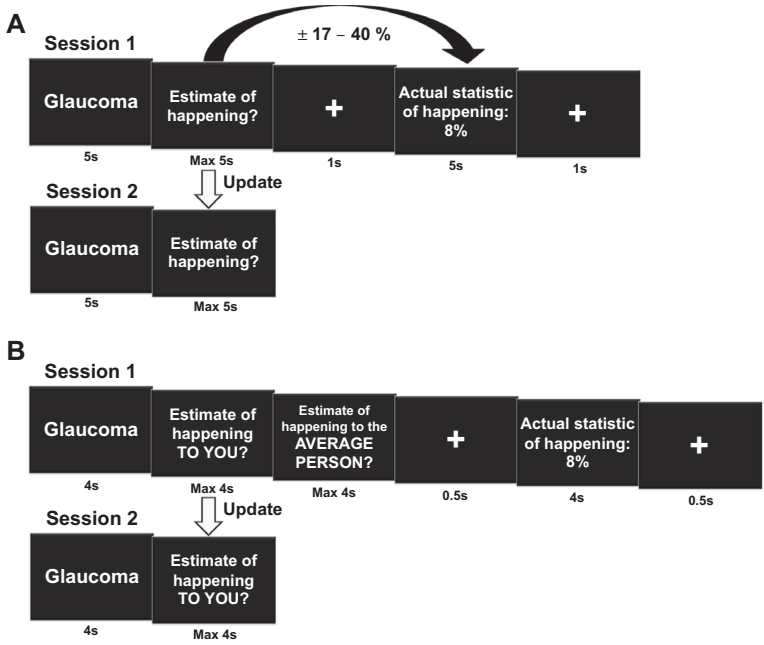

Fig. 1. (A) Procedure for Experiment 1. On each trial participants were provided with one of 80 life events and instructed to estimate the likelihood of that event happening to them. They were then given the 'actual probability' of that event. (B) Procedure for Experiments 2–4. Procedure followed that of Experiment 1 but participants were also asked to provide a base rate estimate.

indicating that the negative event was less likely than had been estimated. On negative undesirable trials a random percentage of the SE1 was added to the SE1, indicating that the negative event was more likely than had been estimated. To illustrate, were a participant to provide an SE1 of 25% in a positive desirable trial, the provided base rate would lie between (25 + .17 × 25=) 29% and (25 + .40 × 25=) 35%. All probabilities were capped between 3% and 77% (as is typical for studies using the update method – e.g., Sharot et al., 2011) and participants were informed that this was the range of possible probabilities.

2.1.3.2. Post experimental tasks. Participants completed two post-experimental tasks immediately after the two sessions.

2.1.3.2.1. Memory errors. Participants were presented with each event again, in a random order, and asked to recall the actual probability of each event. Memory errors were calculated as the absolute difference between the recalled value and the actual probability.

2.1.3.2.2. Salience ratings. Participants were again presented with the events and were asked to rate events on four scales: vividness (How vividly could you imagine this event? From 1 = not vivid to 6 = very vivid); prior experience (Has this event happened to you before? From 1 = never to 6 = very often); arousal (When you imagine this event happening to you how emotionally arousing is the image in your mind? From 1 = not arousing at all to 6 = very arousing); and magnitude of valence (How negative/positive would this event be for you? From 1 = not strongly at all to 6 = very strongly).

2.2. Results

2.2.1. Scoring

For each event the amount of update was calculated first by computing the absolute difference between the SE1 and the SE2, and second, by coding the difference as positive when the update

was in the direction of the base rate and negative when the update was away from the base rate (e.g., if SE1 was 40, base rate was 50, and the SE2 was 30, the update would be coded as −10, whereas if the base rate had been 10 the update would be coded as +10). Thus, negative updates indicate updating in the direction away from the actual base rate. Mean updates for each participant in each condition were then calculated after removal of outliers (±3 × the interquartile range) and trials for which a derived probability could not be applied (e.g., when a participant's SE1 was already at the lowest extreme of the probability range, but the trial-type required that a lower base rate be supplied).

2.2.2. Analysis of updates

Inspection of the mean updates (Fig. 2) revealed an asymmetry in the updating of likelihood estimates with an interaction between the desirability of new information and the type of event (positive or negative) that was judged. For negative events, the typical finding reported using the update method (e.g., Sharot et al., 2011) was replicated. Participants updated their likelihood estimates more in response to desirable than undesirable information. However, for positive events (which have not previously been included in studies using the update method), the pattern of updating was reversed. Participants updated their likelihood estimation more in response to *undesirable* than desirable information. A pattern consistent with the optimism bias account (i.e., greater update in response to desirable information across both negative and positive events) was apparent in just one participant.

Mean updates were entered into a 2 (Event type: positive, negative) × 2 (Desirability: desirable, undesirable information) repeated measures analysis of variance (ANOVA). Neither the main effect of Event type ($F < 1$), nor Desirability ($F < 1$), was significant. However the interaction between these factors was significant, $F(1,12) = 29.39$, $p < .001$, $\eta_p^2 = .71$. Paired sample t-tests indicated that this interaction was due to participants updating significantly more in response to *desirable* than undesirable information when estimating the likelihood of negative events, $t(12) = 4.73$, $p < .001$, $d = 1.70$, but significantly more in response to *undesirable* than desirable information when estimating the probability of positive events, $t(12) = -2.78$, $p = .017$, $d = 1.20$.

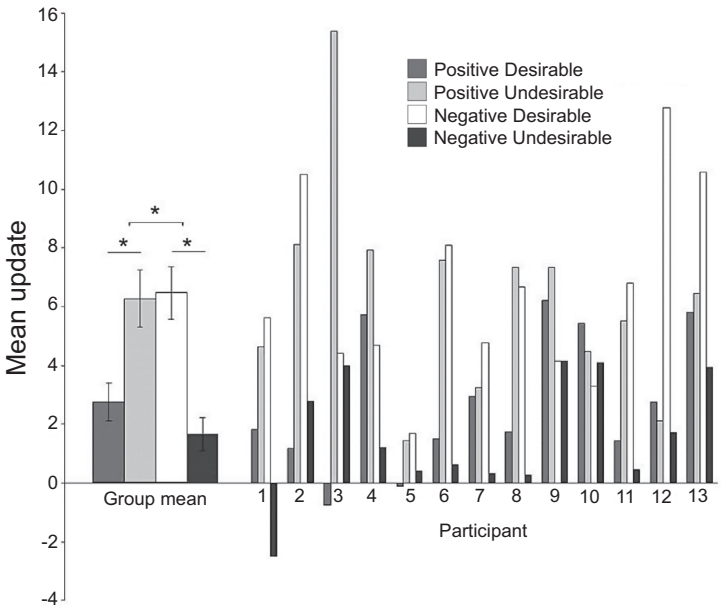

Fig. 2. Mean updates by Event type and Desirability – Experiment 1. Negative updates indicate updating in the direction away from the actual probability. Error bars indicate ±1 standard error of the mean. $^{*}p < .05$, two-tailed.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

So there is an updating asymmetry between desirable and undesirable trials, but it flips across positive and negative events. The replication of a desirability bias for negative events demonstrates that the use of derived probabilities did not influence the way in which participants completed the task. However, the presence of an undesirability bias for positive events is incompatible with an optimism account. This is the major result from Experiment 1. The following analyses simply demonstrate that the result is robust across all potential analyses, including those that control for the possible alternative explanations identified in Sharot et al. (2011), such as differential salience of desirable versus undesirable information.

2.2.3. Additional analyses

2.2.3.1. Uncapped trials. In the typical design of the update method (e.g., Sharot et al., 2011), the number of trials in each cell of the experimental design cannot be controlled because whether the participant receives desirable or undesirable information depends on their SE1. The use of derived probabilities in the present study reduces this problem, but the use of capped probabilities (which were included in order to follow as closely as possible Sharot et al.'s work in all other respects) means that the number of trials per cell may become unbalanced. Since trials are 'capped' at the extreme ends of the distribution, trials are lost when participants enter extreme values. This loss of trials is likely to be unbalanced between conditions. To illustrate, should a participant enter an estimate of 77% (the upper bound of allowed probabilities) then a derived probability on a positive desirable or negative undesirable trial cannot be applied. The same is true for negative desirable and positive undesirable trials when a participant's SE1 is at the lower bound of the capped probability distribution (i.e., 3%). Such extreme trials are excluded from the above analyses (Section 2.2.2) but a less severe problem occurs when SE1s approach the extremes of the probability distribution and the program generates a derived probability that is beyond the cap. The participant is subsequently presented with a 'capped' base rate (either 3% or 77%). It is possible for such trials to be unequally distributed across conditions, and indeed this pattern was observed in the current data, as reflected in a significant interaction between the Desirability and Event type factors in the number of capped trials, $F(1,12) = 4.88$, $p = .047$, $\eta_p^2 = .05$. In order to guard against the differential updating reported above (Section 2.2.2) being due to an unbalanced number of capped trials, a further ANOVA of mean updates excluded *all* capped trials (see Appendix C). This analysis revealed the same pattern of significance as the analysis including capped trials, with a significant interaction between the Desirability and Event type factors, $F(1,12) = 35.33$, $p < .001$, $\eta_p^2 = .75$.

2.2.3.2. Analysis of initial estimates. Inspection of SE1s revealed that participants had a tendency to assign significantly higher initial probabilities to positive events ($M = 35.44$, $SD = 7.28$) than negative events ($M = 29.55$, $SD = 11.62$; $t(12) = 2.48$, $p < .029$, $d = 0.61$). In order to investigate whether the differential pattern of updating reported above (Section 2.2.2) was simply due to the increased probabilities assigned to positive events, a covariate coding for the mean difference in SE1s across desirable and undesirable trials for positive and negative events was entered into the analysis of updates.[2] The interaction between Desirability and Event type remained significant, $F(1,11) = 17.33$, $p = .002$, $\eta_p^2 = .61$.[3] There was still evidence for 'optimistic' updating for negative events, $F(1,11) = 40.47$, $p < .001$, $\eta_p^2 = .79$, although the 'pessimistic' belief updating in response to positive events just failed to attain statistical significance, $F(1,11) = 4.03$, $p = .070$, $\eta_p^2 = .27$.

2.2.3.3. Memory for probabilities. As noted in Sharot et al. (2011), it is possible that differential updating as a function of the desirability of new information is caused by differential memory for desirable and

[2] An ANCOVA is functionally equivalent to including the control variable in a hierarchical regression, since they are both based on the General Linear Model. In the 2 × 2 ANCOVA, the covariate was calculated as follows: (Positive Desirable SE1 – Positive Undesirable SE1) – (Negative Desirable SE1 – Negative Undesirable SE1). In the separate ANCOVAs for positive and negative events, the covariate was simply the calculation from the relevant of the two parenthesized subtractions.

[3] With the present methodology, it is unnecessary to control for the absolute difference between the SE1 and base rate, because this value is directly related to the initial SE1. Consequently, the present analysis controlling for SE1 renders controlling for such a potential difference unnecessary.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

undesirable information. However, participants remembered probabilities equally well (Appendix C) regardless of the valence of the event and desirability of the information. Analysis of memory errors using a 2 (Event type) × 2 (Desirability) repeated measures ANOVA revealed no significant main effects nor interaction (Event type: $F < 1$, Desirability: $F(1,12) = 3.79$, $p = .075$, $\eta_p^2 = .24$, Event type × Desirability interaction: $F < 1$).

2.2.3.4. Analysis accounting for all salience ratings. As a final check of the robustness of the interaction between the Desirability and Event type factors in the analysis of updates, two further analyses were conducted that included all salience ratings (see Appendix C) as covariates (i.e., reported magnitude of event valence, vividness, arousal, and past experience; see Section 2.1.3.2.2). The interaction between Event type and Desirability remained significant whether capped trials were included, $F(1,8) = 15.97$, $p = .004$, $\eta_p^2 = .66$, or not, $F(1,8) = 21.08$, $p = .002$, $\eta_p^2 = .73$, whilst all main effects remained non-significant (all Fs < 1).

2.3. Experiment 1 – discussion

Experiment 1 used a modified version of the update method to investigate the pattern of updating seen in response to desirable and undesirable information when the likelihood of both negative *and* positive events was estimated. It was found that the effect of information desirability on updating was dependent on whether the event being judged was positive (i.e., something a person would want to experience) or negative (i.e., something a person would not want to experience). For negative events, using a paradigm in which it was possible to randomly allocate participants to receive either desirable or undesirable information, we replicated the central finding from the update method (Chowdhury et al., 2014; Garrett & Sharot, 2014; Garrett et al., 2014; Korn et al., 2014; Kuzmanovic et al., 2015, 2016; Sharot, Guitart-Masip, et al., 2012; Sharot, Kanai, et al., 2012; Sharot et al., 2011). Participants updated their personal risk estimates more when provided with desirable information than when provided with undesirable information when judging negative events (and, as in Sharot et al., this result could not be explained in terms of differential event salience, initial probability estimates, or memory). In contrast, participants updated their estimates more in response to undesirable than desirable information when judging positive events. The results of Experiment 1 therefore conflict with a general optimistic pattern of belief updating. They are, however, consistent with previous data suggesting that 'unrealistic optimism' observed in standard, comparison tests of unrealistic optimism is due to statistical artifacts (Harris, 2009; Harris & Hahn, 2011). How then might this result be explained? To this end, the next Section 2.3.1 describes the normative process by which one *should* make estimates of personal risk. We subsequently report a series of simulations (Section 3) that demonstrate how unbiased, completely rational, agents can produce the pattern of results obtained in this experiment. The second set of simulations (Section 3.2) highlights the difficulties associated with methods of measuring update bias, suggesting that no single measure can be used to conclusively test for optimistic belief updating. In the further experimental sections (Sections 4–7) we employ a triangulation approach to determine whether consistent evidence for optimistic updating can be observed across a variety of (individually imperfect) analyses.

2.3.1. Optimism and the logic of risk estimates

Harris and Hahn (2011) showed that there are a number of statistical reasons why the standard comparison method of measuring unrealistic optimism may give rise to seeming unrealistic optimism at a population level even though no individual within that population is optimistic. Rather than focusing on the difference between group and average risk, the update method introduced in Sharot et al. (2011) focused instead on belief change, seeking to detect optimism in the way that people revise their beliefs about risk in response to new information. To illustrate, consider the case in which the national press reports that a deadly virus has found its way into the water supply. National Newspaper A reports that their best estimate is that the virus will affect the water supply of 10% of households in the country. National Newspaper B reports that their best estimate is that the virus will affect the water supply of 30% of households in the country. Amanda reads Newspaper A, and Betty reads Newspaper B. *In the absence of any other information*, Amanda's best estimate of the chance of

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

her house being affected by the outbreak is 10% and Betty's is 30%. If now provided with a new base rate (as in Experiment 1 and Sharot's research) of 20%, Amanda should increase her estimate by 10 percentage points and Betty should decrease her estimate by 10 percentage points, and thus their absolute degree of belief updating should be equivalent. Any systematic difference in the amount of update is evidence of bias.[4]

In general, however, there are *two* distinct ways in which we might receive new information about our personal risk. We may receive new information about the prevalence or base rate (as above), but we may also receive information diagnostic of our own personal risk (e.g., vaccination). Both of these types of information are relevant, both can be desirable or undesirable, and, according to the normative procedure for determining risk, both should be combined via Bayes' Theorem to provide our best estimate of risk (e.g., Hardman, 2009; Kahneman & Tversky, 1973). In a population, some people will have received a vaccination and thus be less at risk, whilst some people will not have received a vaccination and thus be more at risk than the base rate (because the base rate is the average across the entire population). This fact has been recognized by optimism researchers for some time, and is the very thing that makes optimism research so difficult:

> "A woman who says that her risk of heart disease is only 20% ... may be perfectly correct when her family history, diet, exercise, and cholesterol level are taken into consideration, despite the fact that the risk for women in general is much higher".
>
> [Weinstein & Klein, 1996, p. 2]

Weinstein's 'comparative method' was originally designed to overcome specifically this difficulty, though in practice it fails to provide an adequate solution (see Harris & Hahn, 2011). The fact that the normative best estimate of personal risk is a combination of *both* the average person's risk (base rate) and individual diagnostic information also affects the study of belief updating. The following example (Section 2.3.2) outlines how a rational agent should update their risk estimates in light of new information.

2.3.2. Normative risk updating

55-year-old Tim estimates that the average 55-year-old's risk of contracting heart disease (the base rate) is 20%. *In the absence of any other information*, Tim's best estimate of his own likelihood of contracting heart disease is 20%.

If Tim possesses any diagnostic information that differentiates his risk from the average person's, he should normatively combine the base rate with this diagnostic information. For example, if he does not have a family history of heart disease, his risk is lower than the average person's. Bayes' Theorem prescribes how this information should be combined (e.g., Kahneman & Tversky, 1973):

$$P(h|e) = \frac{P(h)P(e|h)}{P(h)P(e|h) + P(\neg h)P(e|\neg h)} \quad (1)$$

Bayes' Theorem prescribes the probability, $P(h|e)$, of experiencing an event h (e.g., heart disease) in light of evidence e (no family history of heart disease). The best estimate of experiencing that event is a function of the base rate of the event, $P(h)$, and the diagnosticity of the evidence – the likelihood ratio, $\frac{P(e|h)}{P(e|\neg h)}$. The likelihood ratio is the ratio between the conditional probability of obtaining the evidence given that the hypothesis is true, $P(e|h)$, and the probability of receiving it when the hypothesis is false, $P(e|\neg h)$. In Tim's case, $P(e|h)$ reflects how likely a heart disease patient is to have no family history of heart disease, whereas $P(e|\neg h)$ reflects the probability of no family history of heart disease in those who do not contract it. From a longitudinal study (Hawe, Talmud, Miller, & Humphries, 2003), we can calculate $P(e|h) = .52$ and $P(e|\neg h) = .66$. As the likelihood ratio is less than one, such evidence (no family history of heart disease) should decrease Tim's estimate of contracting heart disease. Specifically, Tim's estimate of $P(h)$ is 20% and therefore his best estimate of his chance of contracting heart disease combines this with his specific diagnostic information to give:

[4] They may, of course, vary in the magnitude of their original mis-estimate, and this needs to be factored out in statistical analysis.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

$$\frac{.2 \times .52}{.2 \times .52 + .8 \times .66} = 16\% \quad (2)$$

The base rate of heart disease is actually 33% for 55 year-old males (Bleumink et al., 2004). If Tim receives this information, he should recalculate his personal risk once more, using Bayes' Theorem, replacing his previous base rate estimate (20%) with 33%, which will result in an increased 'best estimate':

$$\frac{.33 \times .52}{.33 \times .52 + .67 \times .66} = 28\% \quad (3)$$

Given the two basic components to normative probability judgments – base rates and diagnostic evidence – there are thus two ways to receive undesirable (desirable) new information: One can receive new diagnostic information which suggests that one is more (less) at risk, or one may discover that the base rate is higher (lower) than previously thought. Participants in all studies using the update method (Chowdhury et al., 2014; Garrett & Sharot, 2014; Garrett et al., 2014; Korn et al., 2014; Kuzmanovic et al., 2015, 2016; Moutsiana et al., 2013; Sharot, Guitart-Masip, et al., 2012; Sharot, Kanai, et al., 2012; Sharot et al., 2011) and, following them, participants in the present Experiment 1, did not receive any new diagnostic information. In Eq. (3), Tim knows the accurate base rate, calculates his personal risk rationally, and yet his personal risk is different from the base rate. Individuals should not necessarily change their estimate of personal risk simply because it lies above or below the base rate. Researchers can only discern what effect the new base rate information should have on a participant's risk estimate if they know the participant's previous estimate of the base rate. Without this knowledge, it is impossible to classify a particular trial as 'desirable' or 'undesirable' and therefore it is impossible to say in which direction (and how much) the participant's estimate should change. The following simulation highlights the flaws in the update method by simulating optimistic data from non-biased agents.

3. Simulation

3.1. The problem of misclassification

Take a hypothetical sample of 100 Bayesian agents, 25 of whom assume base rates of .1, .2, .3, and .4, respectively (mean = .25) for Disease X, which has a true base rate of .25. Before the study, these agents receive evidence reflecting their vulnerability to Disease X with the following characteristics: $P(e|h) = .5$; $P(\neg e|\neg h) = .9$. Then $P(h)P(e|h) + P(\neg h)P(e|\neg h)$ defines the proportion who receive evidence suggesting *increased risk*, here: $.25 \times .5 + .75 \times .1 = .2$. Thus 20% of agents have evidence suggesting they will get the disease ('positive evidence') and 80% have evidence suggesting they will not ('negative evidence'). So at each base rate, 5 agents will receive positive evidence, and 20 will receive negative evidence.

In the simulated study (Table 1), agents calculate their initial risk estimates normatively via Bayes' Theorem (Eq. (1)), using their subjective base rates; their second estimate recalculates Bayes' Theorem using the experimenter-provided true base rate (as is exemplified in Eqs. (2) and (3)).[5]

Agents whose subjective base rate estimates were below the true base rate of .25 receive genuinely undesirable information: Disease X is more prevalent than they thought. Agents whose subjective base rate estimates were above the true base rate receive genuinely desirable information: Disease X is less prevalent than they thought. However, Sharot et al.'s (2011) update method classifies infor-

[5] This assumes that the agents perceive the base rate information as maximally reliable. If the agents do not fully trust the 'experimenter' then their new base rate will deviate. In the extreme case in which agents believe the source of the information to be maximally unreliable (i.e., one can infer absolutely nothing from the source's report), one will see no asymmetric updating because one will observe no updating at all. In all other instances, the direction of the asymmetry will remain the same, it is only its magnitude that would change. The only situation in which this would not be the case would be if one built in an assumption whereby the agents trusted information differently according to its desirability. This, however, would of course be a form of bias, and our simulation would no longer be one of rational Bayesian agents, and the whole point of the simulation is to demonstrate how a seemingly biased pattern of results can obtain from unbiased, rational Bayesian agents.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

Table 1
Artifactual unrealistic optimism.

	Those with positive evidence				Those with negative evidence			
Subjective base rate	0.1 ($n = 5$)	0.2 ($n = 5$)	0.3 ($n = 5$)	0.4 ($n = 5$)	0.1 ($n = 20$)	0.2 ($n = 20$)	0.3 ($n = 20$)	0.4 ($n = 20$)
Initial estimate	0.357	0.556	0.682	0.769	0.058	0.122	0.192	0.270
True base rate	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Experimenter-defined desirability	Des	Des	Des	Des	Undes	Undes	Undes	Des
True desirability	Undes	Undes	Des	Des	Undes	Undes	Des	Des
Correctly classified?	NO	NO	YES	YES	YES	YES	NO	YES
Final estimate	0.625	0.625	0.625	0.625	0.156	0.156	0.156	0.156
Amount of update (IE–FE)	−0.268	−0.069	0.057	0.144	−0.098	−0.034	0.036	0.114

Note. Des = desirable information; Undes = undesirable information.

mation as 'desirable' or 'undesirable' based on the relationship between initial estimate and true base rate, thus misclassifying 30% of the sample (gray columns).

On that method, the final experimental 'result' is obtained by averaging across those agents receiving 'desirable' and 'undesirable' information (see 'Experimenter-defined Desirability' in Table 1).

As each positive evidence group represents 5 agents, and each negative group 20 agents, the resulting absolute means are: Desirable Group = |0.04|; Undesirable Group = |−0.03|. Thus, these rational agents show 'greater updating' in response to 'desirable' than 'undesirable' information and would be labelled optimistic – though rational by definition – due to incorrect classification. Although this may seem a somewhat small effect, it becomes much more pronounced when base rate estimates are regressive toward the midpoint of the scale (i.e., values below .5 are overestimated and values above .5 underestimated), as is typical of people's probability estimates in many contexts (see e.g., Harris & Hahn, 2011; Moore & Small, 2008, and references therein). If the true base rate were .21, i.e., below the agents' mean estimate, then the seeming difference in updating rises to 8% ('desirable' = .084; 'undesirable' = .004) – easily sizeable enough to account for extant experimental data (e.g., Experiment 1; Section 2).

Fig. 3 demonstrates that this pattern of results is not dependent on the precise parameters used in this illustrative example. The preponderance of positive differences in updating (where people update more in response to desirable than undesirable information) is clear from both Fig. 3A (where mean estimates of the underlying base rate are correct) and Fig. 3B (where base rate estimates are regressive, and consequently represent more realistic simulations). Note that, were participants required to estimate their chance of *not* experiencing the event (see e.g., Sharot et al., 2011), they would be estimating $P(\neg h)$. For rational Bayesians, all the probabilities would be complements of those in Table 1, and if 'rational' updating appears biased in estimates of $P(h)$, it will also appear biased in estimates of $P(\neg h)$. The direction of the asymmetry 'flips' above and below 50%, but because Sharot et al.'s complement events simultaneously flip the valence of the event, seeming 'optimism' is preserved.

Fig. 3 demonstrates two general characteristics: the fact that the mass of data points are above 0, indicating seeming optimism, and the fact that the landscape contains many sharp boundaries. The fact that these boundaries are so sharp is not only surprising, but also consequential. At these boundaries, a tiny change in either sensitivity, $P(e|h)$, or specificity, $P(\neg e|\neg h)$, of people's diagnostic information leads to significant change in update asymmetry. This explains how a non-selective change to a probabilistically relevant quantity (e.g., changing the perceived diagnosticity of individuals' evidence; or altering the regressiveness of their initial base rate estimates) that affects all agents equally, can lead to a seemingly selective effect: a sharp increase in the difference between updating for 'desirable' vs. 'undesirable' information. Thus, the selective effects of, for example, L-DOPA (Shah, 2012; Sharot, Guitart-Masip, et al., 2012) and transcranial magnetic stimulation (TMS) (Sharot, Kanai, et al., 2012) on belief updating might be entirely unrelated to optimism, and simply reflect (for example) *better* learning (i.e., less conservative updating – formally equivalent to an increase in the diagnosticity of information) following receipt of L-DOPA. Specifically, the *z*-axes of the landscape plots of Fig. 3 represent the *difference* in belief updating between desirable and undesirable trials. Looking at Fig. 3, one

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

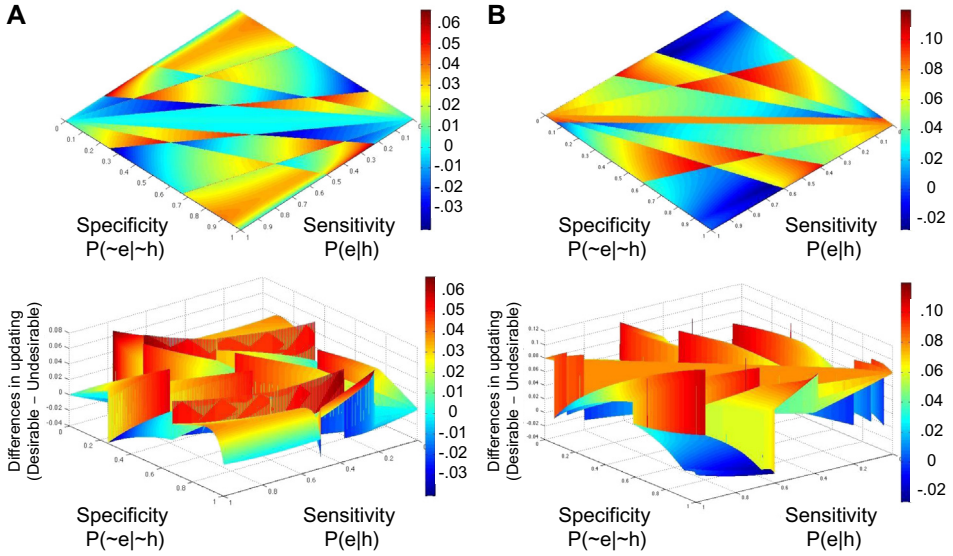

Fig. 3. Seemingly 'optimistic' updating. Positive values in the *z*-axes demonstrate 'optimistic' updating for Bayesian agents receiving evidence with the properties shown in the *x*-axes, $P(e|h)$, and *y*-axes, $P(\neg e|\neg h)$. (A) Agents estimate the base rate as either: .1, .2, .3, .4, and the true base rate = .25 (as in Table 1). (B) True base rate = .21. The mass of data points in both plots spuriously suggests optimistic updating. Both landscapes are shown once from above and once from the front.

can see that this difference (represented on the *z*-axis) is not consistent across the parameter space (*x* and *y*-axes). Any movement within the parameter space will therefore affect the magnitude of this observed difference. Crucially, this movement (reflecting a change in the underlying probabilistic quantities) affects all agents in the simulations equally, thus an experimental manipulation which affects the underlying probabilistic quantities (e.g., how diagnostic an agent considers a given piece of evidence to be) for all agents equally will appear to have a differential impact on desirable and undesirable trials. Consequently, the effect of such a manipulation will be manifest as a statistical interaction between desirability and group effects (e.g., the administration vs non-administration of L-DOPA). The mere fact that the desirability bias can be altered through TMS or other experimental manipulations thus provides no independent evidence that the effect represents a genuinely optimistic asymmetry. Rather, such effects are entirely compatible with the nature of the observed statistical artifact. All that is required to obtain such effects is that the manipulation somehow influences the probabilistically relevant quantities on which all agents base their judgments. A signature of such an *artifactual* interaction is that it should arise in the same direction for positive and negative events, such that both the seeming optimism (for negative events) and seeming pessimism (for positive events) should diminish, or both together should increase. Harris, Shah, Catmur, Bird, and Hahn (2013) show such a result with individuals with Autism Spectrum Disorder.

The critique above seems to suggest a straightforward 'fix' of the updating method, namely the inclusion of participants' estimates of the average person's risk. With this inclusion, enabling the correct definition of desirable and undesirable information, one might expect updating to be equal in response to desirable and undesirable information, unless people were optimistically biased. However, this is not so. The combination of base rate error and diagnostic information presents more fundamental problems: even for correctly classified participants seeming updating asymmetries will ensue. As Appendix D explains, this will be the case even where there is only a single diagnostic test involved (as in Table 1), and these problems are compounded where participants vary in the diagnostic information they possess. This is illustrated in the next section, Section 3.2, which demonstrates why there is no quick 'fix' for the updating paradigm and clarifies the fundamental difficulties inherent in assessing the rationality of belief updating.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

3.2. The problem of diagnostic information and the bounded probability scale

The preceding Section 3.1 outlines the conceptual issues involved in assessing optimistic belief updating. It does not, however, address all measurement issues that are involved in an empirical investigation of whether or not people's belief updating is optimistically biased.

It is clear that two factors determine the beliefs of a rational agent: the base rate (average risk), and whatever individual diagnostic knowledge that agent might possess. In a belief updating experiment such as Experiment 1 or Sharot et al. (2011), the participants are only provided with new information about the *base rate* by the experimenter, but it is participants' revision of their beliefs about personal risk that must be empirically assessed.

If rational agents possessed no diagnostic knowledge and base rate information was all they had to go by, revision should normatively consist of moving to what they perceive the new base rate to be. The amount of belief change will thus simply be the difference between the initial base rate estimate and the revised base rate estimate. Trivially, however, the amount of belief change in rational agents receiving 'desirable' or 'undesirable' information about the base rate will be the same only if the magnitude of their initial estimation error is the same. Any analysis of actual updating behavior must thus seek to control for differences in initial base rate error.

In their seminal paper, Sharot et al. (2011) conducted regression analyses to investigate the degree to which initial errors predicted subsequent update – the term 'learning score' was used for these regressions in Moutsiana et al. (2013; see also Garrett et al., 2014). Such an analysis might seem also to be a way to control for differences in initial error. Sharot et al. regressed the amount of belief change observed (i.e., the difference between first and second self-estimate) on the size of the initial error (in this case defined incorrectly as the difference between initial self-estimate and true base rate). For each participant in the study, two such regressions were conducted, comparing the events for which the participant received 'desirable' versus 'undesirable' information. This yields two coefficients for each participant, and the overall analysis simply compared the coefficients for statistical differences.

If participants had *no* diagnostic knowledge, and their self-estimates depended entirely on their beliefs about the base rate, these regression-based analyses would be appropriate: whatever the size of the initial error, it should be fully reduced on update, making the correlation between 'initial error' and 'belief change' a perfect correlation in a rational agent. However, this is no longer the case if participants believe themselves to be in possession of individual diagnostic knowledge. As outlined in the preceding Section 3.1, diagnostic knowledge means that an individual's self-estimate need no longer equal the base rate. From the perspective of the regression analyses Sharot et al. (2011) conduct, this individual knowledge is simply 'noise' around the underlying base rate estimates. Unfortunately, this 'noise' is unlikely to simply 'cancel out' across conditions. In fact, diagnostic knowledge poses a problem even when 'initial error' is defined relative to the base rate (as is normatively appropriate) and the regression is conducted between belief change concerning self risk and initial base rate error.

More specifically, the root of this problem is that in Bayes' theorem (see above; Section 3.1) individual diagnostic information combines *multiplicatively* with the base rate and is normalized to a bounded scale between 0 and 1 (0% and 100%). This leads to systematic distortion the moment there is variability in diagnostic knowledge across participants/life events.[6] This distortion likely makes a second, independent contribution to Sharot et al.'s (2011) result, over and above the issue of misclassification discussed in Section 2.3.1. This is demonstrated in Table 2, which shows an artificial sample of participants that has been constructed to match exactly both base rate error (BR error) and diagnostic information (LHR – likelihood ratio) across those receiving 'desirable' and 'undesirable' information.

Specifically, there are 28 participants in each 'group'. The 'true' base rate is 30%. For each participant in the 'desirable information' group there is one in the 'undesirable information' group whose *base rate error is exactly the same*, except in sign (under- rather than over-estimating the base rate) and who possesses *the exact same amount of diagnostic knowledge* (represented by the likelihood ratio). The following columns then give, for each participant, their (rounded) estimates of the average per-

[6] Uniform diagnostic knowledge across all life events/persons would be fine as regression is a linear relationship that is unaffected by multiplication of values by a constant (or addition of a constant).

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

Table 2
A perfectly matched sample for an updating experiment in which the true provided base rate is 30%.

	Parameters		Estimates				
	LHR	BR error	BR1	SE1	SE2	SE1 error	Update
Desirable information	1.3	30	60	66	36	36	−30
	1.1	29	59	61	32	31	−29
	0.9	28	58	55	28	25	−27
	0.8	27	57	51	26	21	−25
	0.7	26	56	47	23	17	−24
	0.6	25	55	42	20	12	−22
	0.5	24	54	37	18	7	−19
	0.4	23	53	31	15	1	−16
	0.4	22	52	30	15	0	−15
	0.5	21	51	34	18	4	−16
	0.5	20	50	33	18	3	−15
	0.5	19	49	32	18	2	−14
	0.6	18	48	36	20	6	−16
	0.7	17	47	38	23	8	−15
	0.8	16	46	41	26	11	−15
	0.55	15	45	31	19	1	−12
	0.55	14	44	30	19	0	−11
	0.6	13	43	31	20	1	−11
	0.6	12	42	30	20	0	−10
	0.65	11	41	31	22	1	−9
	0.75	10	40	33	24	3	−9
	0.7	9	39	31	23	1	−8
	0.7	8	38	30	23	0	−7
	0.8	7	37	32	26	2	−6
	0.8	6	36	31	26	1	−5
	1.2	5	35	39	34	9	−5
Undesirable information	1.3	−30	1	1	36	−29	35
	1.1	−29	1	1	32	−29	31
	0.9	−28	2	2	28	−28	26
	0.8	−27	3	2	26	−28	24
	0.7	−26	4	3	23	−27	20
	0.6	−25	5	3	20	−27	17
	0.5	−24	6	3	18	−27	15
	0.4	−23	7	3	15	−27	12
	0.4	−22	8	3	15	−27	12
	0.5	−21	9	5	18	−25	13
	0.5	−20	10	5	18	−25	13
	0.5	−19	11	6	18	−24	12
	0.6	−18	12	8	20	−22	12
	0.7	−17	13	9	23	−21	14
	0.8	−16	14	12	26	−18	14
	0.55	−15	15	9	19	−21	10
	0.55	−14	16	9	19	−21	10
	0.6	−13	17	11	20	−19	9
	0.6	−12	18	12	20	−18	8
	0.65	−11	19	13	22	−17	9
	0.75	−10	20	16	24	−14	8
	0.7	−9	21	16	23	−14	7
	0.7	−8	22	16	23	−14	7
	0.8	−7	23	19	26	−11	7
	0.8	−6	24	20	26	−10	6
	1.2	−5	25	29	34	−1	5

Note. LHR = likelihood ratio, BR error = base rate error, BR1 = base rate estimate, SE1 = initial self estimate, SE2 = second self estimate, SE1 error = difference between SE1 and true base rate.

son's risk (base rate, BR1), their initial self estimate (SE1), their revised self estimate (SE2), and the amount of belief change (Update; SE2–SE1). All participants' estimates are normatively derived via Bayes' theorem.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

Despite the fact that this hypothetical sample of participants is perfectly balanced (far more than any real sample could ever hope to be) there are differences both in the amount of absolute belief change and in the correlation between initial base rate error and belief change. The correlation between initial base rate error and belief change for the 'desirable information' participants is .96, but for the 'undesirable information' participants only .88 – even though all participants are *updating fully, in a normatively correct manner, and initial error and diagnostic knowledge are perfectly matched.*

The situation is even worse if, following Sharot et al. (2011), the correlation is calculated between 'SE1 error' (the deviation between SE1 and true base rate) and belief change. In this case the correlation for the 'desirable information' participants is .86, yet for the 'undesirable information' participants it is .74. And this is without the additional problem of misclassification. In this hypothetical sample, all participants have been classified appropriately relative to base rate error.

How can these discrepancies arise? And why are they more severe if the correlation is calculated with the SE1 error as the reference point? The basis for this regression artifact lies in the compressed nature of the probability scale illustrated in Fig. 4A and B. Correcting base rate error means moving through this scale. However, those moving upwards to increase their estimates (on receiving 'undesirable information') are moving through a different part of the scale than those moving down.

To illustrate with an example: where the true base rate equals 30%, someone over-estimating that base rate would, in the absence of diagnostic knowledge, move from 45% toward 30% on receiving the desirable information, whereas someone who had underestimated by the same amount would move from 15% to 30%. However, once both individuals are in possession of diagnostic knowledge (for example they are less at risk, e.g., the likelihood ratio = 0.5) they would move from 29% to 17.6%, and 8% to 14% respectively. That is, the person in receipt of desirable information would (normatively) have to move 12.4 percentage points, but the person receiving undesirable information would have to move only 6 percentage points. The regression analysis, however, cannot 'know' this, because it does not factor in diagnostic knowledge, and expects equal amounts of belief change from both.

The problem is aggravated where, as in Sharot et al.'s analyses, the calculation of initial error is based on the self-estimate and not the base rate. Equal distance from the true base rate when measured from self-estimates implies *greater* diagnostic knowledge for those below than above

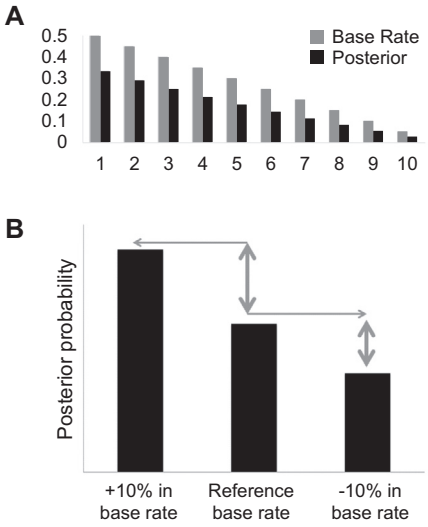

Fig. 4. Probability scale compression. Panel A shows, along the *x*-axis, 10 different pairs of base rate and posterior probability (*y*-axis). Across all 10 pairs the likelihood ratio, that is, the degree of individual diagnostic knowledge is the same, and the posterior probability is derived, via Bayes' theorem, by combining that diagnostic knowledge with the respective base rate. Panel B highlights that a 10% difference in base rate above or below a reference base rate corresponds to different % changes in posterior degree of belief.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

(see Fig. 4B), if base rate error is held constant. In other words, for a self-estimate to be equally far 'below' the true base rate – given scale compression – one needs to be comparatively *even less* at risk relative to the average person. That is, one needs to possess even more diagnostic knowledge indicating lower risk, and as a consequence, one should (normatively) exhibit even less belief updating. Normatively, belief change depends on the multiplication of individual diagnostic information (the likelihood ratio) with base rate information. Hence diagnostic information weights the impact of the base rate: the more diagnostic the individual information is, the less influential the base rate is for the aggregate judgment (e.g., a man without a bike cannot have it stolen no matter what the base rate is), and vice versa.

So for two rational agents whose SE1s are the same absolute distance above and below the base rate, the agent whose estimate lies below *must*, normatively, revise less on receipt of the new base rate, because a smaller proportion of that agent's individual risk derives from the base rate in the first place, as long as base rate error is the same, and both agents are in receipt of diagnostic information indicating less than average risk (as the majority of agents, in the case of events with frequency below 50%, will be). In other words, equating these agents whose initial self-estimates lie equally far above and below the base rate brings about 'asymmetric updating' by mathematical necessity, if these agents are matched on base rate error. The same is true if the diagnosticity of their individuating knowledge is held constant, and base rate error is allowed to vary. It is simply impossible to match participants' individual diagnostic information, participants' base rate error and the amount by which they have to update their beliefs. Only two of these three can be simultaneously matched across those receiving desirable and undesirable information. Likewise, if one matches agents above and below the base rate on their initial deviation from that base rate (SE1 Error), they must differ in either diagnosticity, base rate error, or both. This can be seen by examining Table 2.

The majority of participants receive diagnostic information that indicates they are less at risk than the average person, that is, they have likelihood ratios below 1 (give the base rate of 30%, 70% will not go on to experience the event and the distribution of diagnostic knowledge must reflect that). Picking out any two matched desirable/undesirable information participants shows the above effects. Compare, for example, the third participant in each condition of Table 2: these two participants, by design, are matched in likelihood ratio (LHR) and base rate error (BR Error). The 'undesirable information' participant has greater SE1 Error than the 'desirable information' participant. In order to bring the 'undesirable information' participant's error in line with that of his desirable information counterpart, either his base rate error has to be reduced or his likelihood ratio has to be made smaller (or both). Either of these changes will mean that the 'revised', now matched in SE1 Error participant, should update less on receiving the new base rate estimate.[7]

This affects not only the magnitude of absolute change but also correlations controlling for initial error. And it affects not only conceptually problematic correlations between SE1 and update as conducted by Sharot et al. (2011). It also affects more sensible attempts to factor out differences in initial base rate error by correlating initial base rate error with update.

This is the problem underlying the perfectly matched sample of Table 2. The pernicious effect of individual diagnostic information can be further illustrated with respect to this table: simply replacing the four most extreme likelihood ratios in each group with a less diagnostic value (specifically, replacing .4, .5, .5, and .5 with .7 in each group) reduces the difference in correlation between the two groups from .12 to .09 (on Sharot's correlation between SE1 and update) and from .08 to .06 (on the correlation between initial base rate error and update).

This change illustrates also that the exact outcome of the regressions for participants receiving desirable vs. undesirable information will depend on the exact degrees of diagnostic knowledge people possess, even where that diagnostic knowledge and base rate error, and the pairing of diagnostic knowledge and base rate error, have been exactly matched as in Table 2. The regression outcomes will vary all the more with real samples, where base rate error, the degree of diagnostic knowledge, and its distribution are all subject to sampling variability.

[7] The reverse happens where the LHR is greater than 1, that is, participants are in receipt of information indicating they are at greater risk. The description also does not hold for those participants whose diagnostic information and base rate error is such that the desirability of their new information would have been misclassified (see Section 3.1).

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

The artifactual differences in the regression between groups are thus themselves necessarily stochastic and subject to variation. Specific combinations can generate both seeming 'optimism' and seeming 'pessimism' from entirely rational, fully updating, agents. While it is thus clear that the regression analyses conducted in the majority of studies utilizing the update method (e.g. Sharot et al., 2011) are statistically inappropriate, and their results are neither interpretable nor meaningful with respect to the question of bias in human belief revision, one may still ask to what extent regression artifacts will generate data patterns such as those Sharot et al. observed, beyond the confines of carefully controlled hypothetical data sets such as Table 2. One may ask further to what extent these artifacts persist even on the more sensible correlation between base rate error and update.

To this end we conducted Monte Carlo simulations of samples of rational, Bayesian agents who update equally (namely fully) in receipt of desirable and undesirable information about the base rate. Base rate error and 'diagnostic knowledge' were randomly generated and participants were classified as receiving 'desirable information' or 'undesirable information'. Specifically, the simulations allow one to specify sensitivity and specificity of a diagnostic 'test', add Gaussian random noise in order to generate variability across individuals, and to generate numbers of those receiving test results indicating greater and lesser than average risk in accordance with the base rate (see Harris & Hahn, 2011). We then evaluated the number of participants that are classified into both the 'desirable information' and the 'undesirable information' condition on both the normatively appropriate base rate classification scheme and classification based on initial self-estimate. We then evaluated differences in initial error and absolute belief change, and the resulting correlation between 'initial error' (on both measures of initial error: that is, deviation from self estimate, and deviation from base rate estimate) and belief change.

Figs. 5 and 6 show sample runs of these simulations. Fig. 5 plots the results for agents with *no* diagnostic knowledge. In this case, all final self-estimates fall perfectly on the diagonal, indicating that these agents have updated as much as they should on the basis of initial error (which in this case consists only of base rate error by definition). Fig. 6 shows what happens to those same plots when diagnostic knowledge is incorporated into the update. First, the simulation replicates the main findings

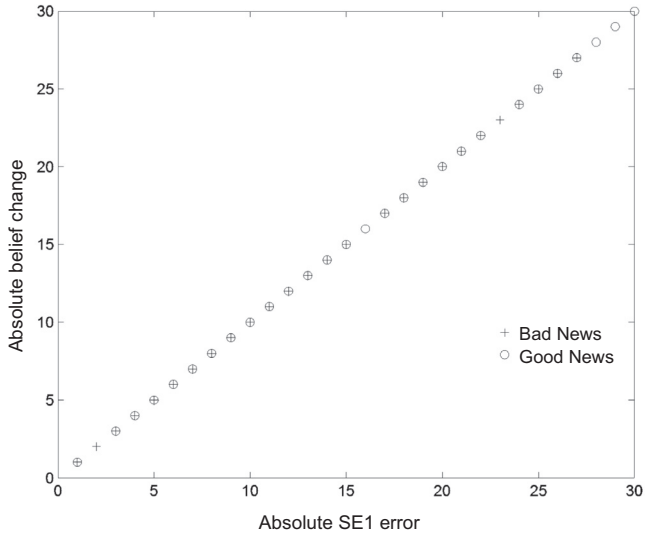

Fig. 5. Simulated rational participants with no individual diagnostic knowledge. The figure plots the (absolute) 'error' in initial self estimate (SE1) as calculated by Sharot and colleagues (i.e., initial self-estimate – 'true' base rate) against (absolute) change, that is, the difference between initial and final self-estimate. Separate markers represent the participants who received desirable information upon hearing the true base rate, and those for whom this was undesirable information. As participants are fully rational, all data points fall on the diagonal, giving rise to a perfect correlation between initial error and amount of change.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

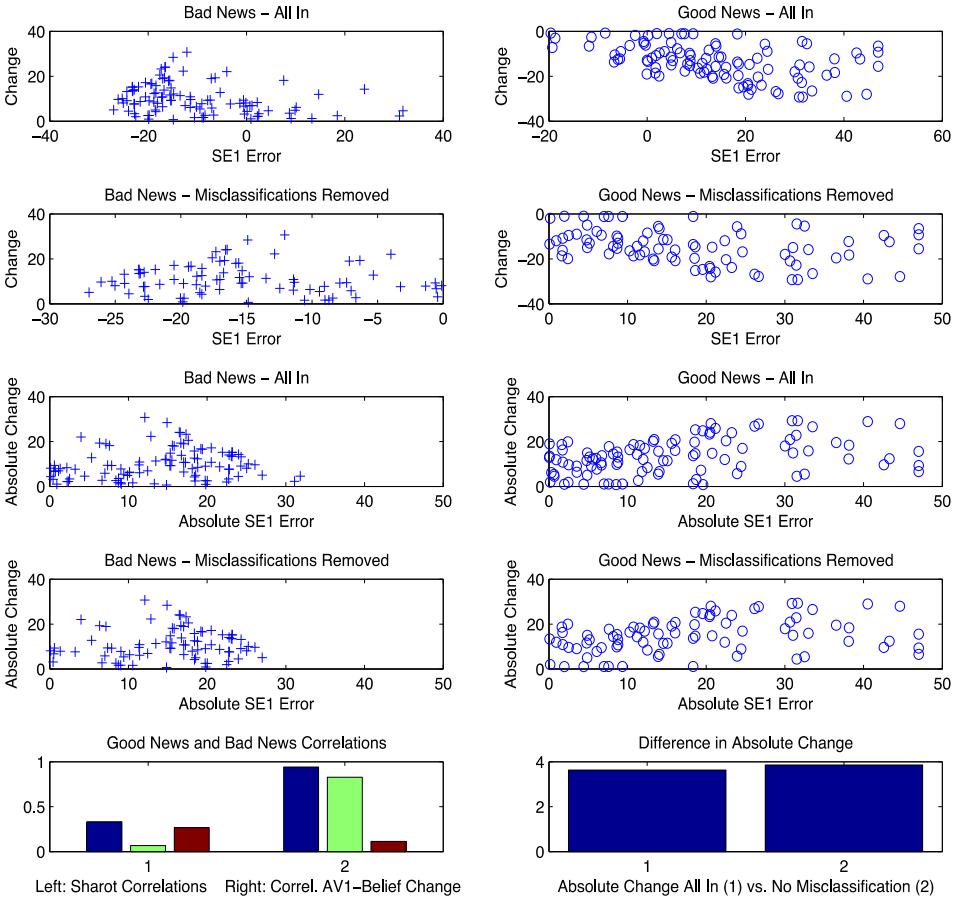

Fig. 6. Rational participants with diagnostic knowledge. The figure shows the impact of providing the same rational agents shown in Fig. 5 with individual diagnostic knowledge. The top four rows show plots of belief update against error in initial self estimate (SE1) as calculated by Sharot et al. (2011), that is, the difference between that self estimate and the true base rate. This relationship forms the basis of Sharot et al.'s controls for differences in individual error via comparison of correlations for 'desirable information' and 'undesirable information' trials. In the bottom left panel, Column 1 shows Sharot correlations resulting from correlation between absolute belief update and SE1 error for the simulated data shown in the panels above, with blue = desirable information, green = undesirable information, and red = the difference between desirable and undesirable information correlations. Note that the correlations shown are for the correctly classified participants only (rows 2 and 4). In the bottom left panel, Column 2 shows the corresponding correlations between absolute belief update and initial base rate error. The bottom right panel shows asymmetrically optimistic updating in terms of absolute updates. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

from the update method (e.g., Sharot et al., 2011). In line with 'optimistic updating' the 'desirable information' people show significantly greater absolute update than the 'undesirable information' people. Furthermore, those in receipt of 'desirable information' (indicated by circles, right hand side of plot) show a greater correlation between initial error (relative to their initial self estimate as a reference point) and their update than those in receipt of 'undesirable information' (left hand side of the plot). All of this occurs even though all agents in the simulation are, by design, rational and unbiased. We next take the reader through Fig. 6 in more detail.

As a reminder, the normatively appropriate, correct, classification scheme classifies participants as receiving desirable or undesirable information about the base rate from the experimenter according to their initial base rate estimates, not according to whether or not the experimenter provided, 'true' base rate is above or below their own initial self-estimate (SE1) as in the original update method (e.g.,

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

Sharot et al., 2011, which we intentionally followed in Experiment 1). The top four rows of Fig. 6 contain plots both of the simulated participant data with all participants in the sample (with $N = 100$ for both 'desirable information' and 'undesirable information' conditions), and plots with only correctly classified participants ('desirable information' = 82, 'undesirable information' = 80), in order to give a feel for the overall distributions and how they are affected by misclassification (as already discussed in Section 3.1).

Fig. 6 (bottom right panel) then shows clearly that the asymmetric updating effects are not solely due to participants who are misclassified on the Sharot classification scheme discussed above. The asymmetry in absolute change for those in receipt of desirable vs. those receiving undesirable information (bottom right panel) is present both for all participants, and among the subset who are classified correctly.[8]

The crucial panel of Fig. 6, however, is the bottom left panel. Sharot et al. (2011) demonstrate a difference in correlation between initial and final estimates for 'desirable information' and 'undesirable information' participants. The bottom left panel ('Sharot scheme') shows that the simulated data replicate the differences in correlation between the two groups.

The correlation values plotted are based on correctly classified participants only. This demonstrates clearly that the regression artifacts created by the failure to control appropriately for individual diagnostic knowledge provide a separate, independent basis for the seeming optimistic updating effect. This also renders moot Garrett and Sharot's (2014; see also Kuzmanovic et al., 2015) demonstration (using negative life events) that updating asymmetries arise even where participants are correctly classified.

Finally, Fig. 6 bottom left panel also shows why it is better to control for initial error by correlating belief change with initial base rate error. In this particular simulation, the difference between 'desirable information' and 'undesirable information' correlations is halved. However, this does not fully solve the problem: while correlations for 'desirable information' and 'undesirable information' people will typically be more similar, as more meaningful quantities are being correlated (this is reflected also in the correlations being higher overall), differences, and even statistically significant differences, can remain.

Both types of 'control' for initial error are problematic because they do not factor in appropriately the effects of individual diagnostic knowledge, and both can thus lead to seemingly optimistic updating. However, Monte Carlo simulations like the ones just described suggest that the differences in correlation are typically much less marked when 'initial error' is defined appropriately, as is in fact observed in Garrett and Sharot (2014, Fig. 3B, left panel).

Finally, it should be noted that the regression artifact need not always lead to seeming optimism. Depending on the underlying characteristics of the diagnostic 'test' (that is, its sensitivity and specificity) and the base rate, other patterns can also ensue. There is thus not even an expectation that Sharot et al.'s finding need replicate with other events.[9]

All of this raises the question of how one might test appropriately for optimistic belief updating. Performing the regressions relative to the correct 'initial error' is better but itself imperfect. The only fully appropriate solutions require factoring in diagnostic knowledge.

The simple addition of a base rate estimate (as already required for appropriate error calculation and trial classification) does, however, allow one to estimate this diagnostic knowledge without explicitly asking participants to provide yet another estimate. Specifically, one can calculate an *implied likelihood ratio* from the base rate estimate and the initial self-estimate. Because Bayes' theorem can be

[8] Note that this sample matches Sharot's findings even in the fact that difference in absolute belief change between the two groups is statistically significant, even though there is no significant difference in the 'error' (deviation from the base rate) of the initial self-estimates (SE1 error).

[9] For those interested in replicating these simulations, the parameters underlying this particular run were 'true', base rate was 30%, base rate error was slightly regressive overall at a mean estimated base rate (AV1) = 32%, sensitivity = .52, specificity, = .65, with Gaussian noise with $m = 0$, sigma = .15. Finally, the script implemented Sharot et al.'s (2011) procedure of capping participant responses to the range 3–77%. Neither regressive base rates, nor capping are necessary for data patterns such as these. However, regressive base rate estimates seem to broaden the range of values for sensitivity and specificity under which the Sharot-style data patterns are observed. The impact of capping, finally, varies with the base rate in question and how diagnostic individuating information is.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

reversed (see Section 4.2.4), one can calculate what likelihood ratio would (normatively) have to be present in order to arrive at the self-estimate provided, given the base rate the participant has estimated. This likelihood ratio can then be combined with the new, 'true', base rate to derive a *predicted* revised self-estimate, which can be compared with the actual revised estimate obtained.

Such a comparison is normatively appropriate. It should be acknowledged, however, that it may suffer from potentially uneven effects of response noise, for the same reasons that diagnostic information has uneven effects. Participants may misremember or mis-select output values, for example mistakenly typing "29" instead of "28". However, a constant absolute amount of noise on the estimate corresponds to a different proportion depending on where one is on the scale. This becomes apparent simply by reversing the logic of Fig. 4: constant differences between values correspond to ever-increasing differences in proportion as one approaches the end points of the scale. Moreover, noise on response estimates arises not only through the pressures of the task, but is a feature of the fixed resolution of the response scale. In rating tasks such as Sharot et al. (2011), participants are not free to respond with any number they wish, but rather are limited to integer responses even if they were able to resolve probabilities differing only in subsequent decimal points. Yet in terms of relative risk, the difference between 3.2% and 1.5% corresponds to the same 10 percentage point drop in base rate as does the difference between 23% and 19.7% for individuals in possession of exactly the same diagnostic knowledge. And rounding has a correspondingly larger effect for the former than for the latter.

By the same token, a failure to update fully (as is, in fact, to be expected on the basis of the large literature on 'conservatism' in human belief revision, e.g., Edwards, 1968; Phillips & Edwards, 1966), may lead to different assessments when considered in terms of absolute differences in update (across desirable and undesirable information trials) or in terms of ratios or proportions (i.e., actual belief change as a proportion of the normatively mandated/predicted belief change).

In the studies described below (Sections 4–7), our strategy, in light of the various difficulties outlined thus far, will be to report the results of a range of possible analyses, across both positive and negative events, to fully test for the existence of optimistic belief updating. If there really is a genuine optimistic bias (certainly an optimistic bias that could have any practical relevance) it should emerge in consistent patterns across these various measures. Perhaps the strongest evidence for optimism would be observed should the comparisons with Bayesian predictions consistently provide evidence for optimistic belief updating. However, as outlined above, this should be consistent across both the ratio measure and a difference measure, as either alone is compromised by different aspects of the bounded probability scale. Table 3 provides a summary of results across all experiments reported in this paper. We contend that it would take a very selective form of motivated reasoning to conclude from these results that people typically update their risk estimates in an optimistic fashion. In all but one study, the central 'result' reported from previous studies using the update method is observed: seeming optimistic updating with negative events using the personal risk classification scheme. This indicates that the different minor changes in method utilized across these studies do not compromise that result. Crucially, however, once positive events are included, there is no consistent evidence for optimism. For none of the studies does the pattern of results, across negative and positive events, display any consistency across these event types or methods of analysis. The methodologies and results of these experiments are provided in further detail in Sections 4–7. Readers less interested in the experimental details of these further studies are encouraged to skip to Section 7.3.

4. Experiment 2

Experiment 2 provided a replication of Experiment 1, but additionally elicited participants' estimates of the population base rate of each event (see Appendix B, Table B2, for mean estimates at each stage). This enabled us to perform the analyses described in Section 3.

4.1. Method

4.1.1. Participants

Seventeen healthy participants (9 females; aged 18–44 [median = 23]) were recruited via the Birkbeck Psychology participant database. All gave informed consent and were paid for their participation.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

Table 3
Results summary – Experiments 1–4.

	Personal risk classification		Base rate classification[a]		Significant difference between classification schemes?	Comparison with rational Bayesian predictions			
						Ratio measure		Difference measure	
	Negative events	Positive events	Negative events	Positive events		Negative events	Positive events	Negative events	Positive events
Experiment 1	Optimism	Pessimism	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Experiment 2	Optimism	Pessimism	n.s.	n.s.	Yes	Optimism	Pessimism	Optimism	Pessimism
Experiment 3A	Optimism	Pessimism	n.s.	n.s.	Yes	n.s.	n.s.	n.s.	n.s.
Experiment 3B	Optimism	Pessimism	n.s.	Pessimism	Yes	n.s.	Pessimism	n.s.	n.s.
Experiment 4	n.s.	Pessimism	n.s.	n.s.	No (p = .08)	n.s.	n.s.	n.s.	n.s.

[a] *Note*. Results are shown after controlling for initial error in base rate estimates.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

4.1.2. Stimuli

Eighty short descriptions of life events, the majority of which had been used in Experiment 1, were presented in a random order. Again, half of the events were positive and half negative. The stimulus set (Appendix E) was slightly altered from Experiment 1 in order to better equate SE1s (participants' initial estimates of their own risk) for positive and negative events, and to reduce the number of illness-related events. Very rare or very common events were again avoided.

4.1.3. Procedure

The procedure followed that of Experiment 1, except that participants were asked to provide a base rate estimate (BR1) after estimating the probability of personally experiencing the event (see Fig. 1B). All base rates were capped between 3% and 80% and participants were informed that this was the range of possible probabilities. Participant feedback suggested that the 5 s presentation times in Experiment 1 were excessively long and therefore presentation times were reduced to 4 s in Experiment 2. The funneled debriefing, as implemented in Experiment 1, revealed that no participants suspected that event base rates were derived from their SE1s.

4.2. Results and discussion

Our first analysis sought to probe broadly whether the differences between the two classification schemes, the scheme used in Sharot et al. (2011), and the normatively appropriate scheme, were empirically consequential by examining whether they led to significant differences in results. As a reminder, Sharot et al. (2011) classified trials as involving 'desirable'/'undesirable' information relative to the participant's initial self-estimate (henceforth referred to as 'personal risk classification scheme'), whereas the normatively correct classification scheme bases classification on the relationship between 'actual base rate' and the participant's base rate estimate (henceforth 'base rate classification scheme').

4.2.1. Comparison of the two classification schemes

A 3-way ANOVA across the two classification schemes demonstrated a 3-way interaction between Classification Scheme, Event type and Desirability, $F(1,16) = 11.42$, $p = .004$, $\eta_p^2 = .42$. The Event type $\times$ Desirability interaction was significantly smaller under the normatively appropriate base-rate classification scheme than under the inappropriate personal risk classification scheme.

This will result from the fact that the use of the personal risk classification scheme will lead to misclassification of trials (see Section 3.1). This was investigated further by analyzing the number of trials in each of the four cells of the design. While the personal risk classification scheme results, by experimental design, in an even distribution across the four trial types (Fig. 7A), numbers vary under the base rate classification scheme (Fig. 7B). When number of trials was entered into a Classification Scheme $\times$ Event type $\times$ Desirability repeated measures ANOVA, the 3-way interaction between the factors was significant, $F(1,16) = 49.95$, $p < .001$, $\eta_p^2 = .76$, suggesting that a significant proportion of datapoints are misclassified on the personal risk classification scheme (as in the simulation with rational agents, Table 1). Indeed, per participant, an average of 12/40 negative events and 8/40 positive events were classified differently under the two classification schemes.

We next analyzed the results in more detail, starting with the normatively inappropriate personal risk classification scheme (the method used in Experiment 1 and by Sharot and colleagues). This analysis tests the replicability of the pattern of results observed in Experiment 1, and checks that the inclusion of questions about the events' base rates did not alter the way in which participants estimated their personal risk. As in Experiment 1, additional analyses including covariates coding for differences in SE1s across positive and negative events, and for salience ratings, were conducted and produced the same patterns of significance, as did analyses including only uncapped trials. These additional analyses are therefore not reported further.

4.2.2. Analysis: personal risk classification scheme

When trials are classified according to the personal risk classification scheme, inspection of the mean updates revealed the same pattern of updating seen in Experiment 1 (Fig. 8A), which was pre-

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

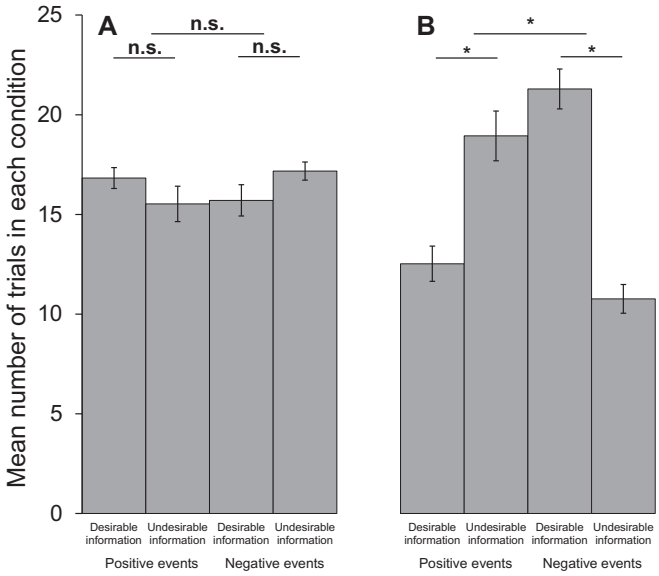

Fig. 7. Mean number of trials by Event type and Desirability – Experiment 2. (A) Normatively inappropriate personal risk classification scheme. (B) Normatively appropriate base rate classification scheme. Error bars indicate ±1 standard error of the mean. $^{*}p < .05$, two-tailed.

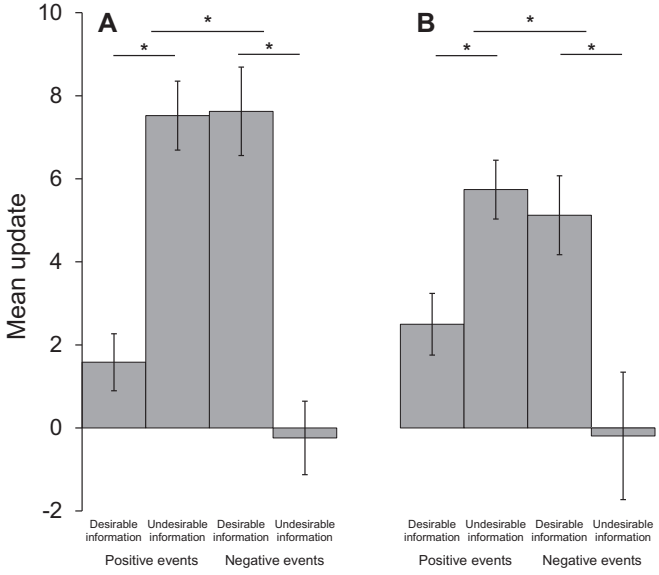

Fig. 8. Mean updates by Event type and Desirability – Experiment 2. (A) Normatively inappropriate personal risk classification scheme. (B) Normatively appropriate base rate classification scheme. No difference in updating was observed under the base rate classification scheme after the degree of base rate estimation error was accounted for. Negative updates indicate updating in the direction away from the actual probability. Error bars indicate ±1 standard error of the mean. $^{*}p < .05$, two-tailed.

sent in all participants. Average updates were entered into a 2 (Event type: positive, negative) × 2 (Desirability: desirable, undesirable information) repeated measures ANOVA. Neither the main effect of Event type, $F(1,16) = 1.59$, $p = .23$, $\eta_p^2 = .09$, nor that of Desirability, $F(1,16) = 2.21$, $p = .16$, $\eta_p^2 = .12$, was significant.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

Once again, there is a flip in asymmetric updating across negative and positive events: the interaction between the factors was significant, $F(1,16) = 54.00$, $p < .001$, $\eta_p^2 = .77$. This interaction was due to participants updating significantly more in response to desirable than undesirable information when estimating the probability of negative events, $t(16) = 6.44$, $p < .001$, $d = 1.93$, but significantly more in response to undesirable than desirable information when estimating the likelihood of positive events, $t(16) = 5.64$, $p < .001$, $d = 1.88$. All these results were unchanged when controlling for differences in SE1s across conditions.

Experiments 1 and 2 therefore gave equivalent results when analyzed using the personal risk classification scheme. This replicates the standard findings observed with negative events (Chowdhury et al., 2014; Garrett & Sharot, 2014; Garrett et al., 2014; Korn et al., 2014; Kuzmanovic et al., 2015, 2016; Sharot, Guitart-Masip, et al., 2012; Sharot, Kanai, et al., 2012; Sharot et al., 2011); but the opposite pattern of results observed in response to positive events means that a general optimistic bias cannot explain these results. Furthermore, these results demonstrate that requiring participants to provide an estimate of base rates in this experiment did not affect participants' pattern of belief updating.

4.2.3. Analysis: base rate classification scheme

When trials are classified using the normatively appropriate base rate classification scheme, the same pattern of differential updating seen in Experiment 1 is evident, but its magnitude is reduced with respect to that observed using the personal risk classification scheme (Fig. 8B), and it is present in fewer participants (65%).

Average updates were entered into a 2 (Event type) × 2 (Desirability) repeated measures ANOVA. Neither the main effect of Event type, $F(1,16) = 2.92$, $p = .11$, $\eta_p^2 = .15$, nor that of Desirability, $F(1,16) = 1.75$, $p = .21$, $\eta_p^2 = .10$, was significant. The interaction between the factors was significant, $F(1,16) = 20.52$, $p < .001$, $\eta_p^2 = .56$, reflecting significantly greater updating in response to desirable than undesirable information when estimating the probability of negative events, $t(16) = 3.45$, $p = .003$, $d = 0.96$, but significantly greater updating in response to undesirable than desirable information when estimating the likelihood of positive events, $t(16) = 4.05$, $p = .001$, $d = 1.08$.

Although it is not a perfect test of optimistic belief updating (see Section 3.2), it is still of interest to see if any evidence of selective updating remains once initial estimation error has been controlled for. A covariate controlling for initial base rate estimation error was therefore included in the analysis.[10] In this ANCOVA, neither the main effects of Event type, $F(1,15) = 2.15$, $p = .16$, $\eta_p^2 = .13$ or Desirability ($F < 1$), nor the interaction between these factors remained significant, $F(1,15) = 1.85$, $p = .19$, $\eta_p^2 = .11$. Analyses of simple main effects revealed that the effect of Desirability was not significant when judging either positive or negative events, $F(1,15) = 1.70$, $p = .21$, $\eta_p^2 = .10$; $F(1,15) = 1.15$, $p = .30$, $\eta_p^2 = .07$, respectively.

4.2.4. Analysis: a comparison with rational Bayesian predictions

As mentioned at the end of Section 3, the addition of a base rate estimate enables the calculation of an implied likelihood ratio:

$$Posterior\ Odds = Prior\ Odds \times LHR \tag{4}$$

$$\Rightarrow \frac{P(h|e)}{1 - P(h|e)} = \frac{P(h)}{1 - P(h)} \times LHR \tag{5}$$

$$\Rightarrow LHR = \frac{P(h|e)}{1 - P(h|e)} \div \frac{P(h)}{1 - P(h)} \tag{6}$$

[10] An ANCOVA is functionally equivalent to including the control variable in a hierarchical regression, since they are both based on the General Linear Model. In the 2 × 2 ANCOVA, the covariate was calculated as follows: (Positive Desirable base rate estimate error − Positive Undesirable base rate estimate error) − (Negative Desirable base rate estimate error − Negative Undesirable base rate estimate error). In the separate ANCOVAs for positive and negative events, the covariate was simply the calculation from the relevant of the two parenthesized subtractions.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

In the terminology of the current experiments, once the initial base rate estimate (BR1) and SE1 are divided by 100, Eq. (6) can be written as:

$$LHR = \frac{SE1}{1 - SE1} \div \frac{BR1}{1 - BR1} \tag{7}$$

Knowing the diagnosticity of the evidence that participants believe they possess (which allows them to differentiate their personal risk from the average person's risk), one can calculate predicted values of SE2, by using the provided base rate information and the calculated LHR to obtain the posterior odds in Eq. (4). A predicted posterior is then obtained through dividing the posterior odds by (1 + posterior odds). As an example, consider an individual estimated their own risk of contracting lung cancer as 10% (SE1) and the average risk as 15% (BR1). Using Eq. (7), these responses imply a likelihood ratio of: $\frac{0.1}{1-0.1} \div \frac{0.15}{1-0.15} = 0.63$ The individual then learns that the base rate is actually 20%. From Eq. (5), their predicted posterior odds are therefore: $\frac{0.2}{1-0.2} \times 0.63 = 0.16$, and therefore the predicted posterior is: $\frac{0.16}{1+0.16} = 0.14$, or 14%.

Optimistic belief updating can then be tested for with either a difference measure (predicted belief change − observed belief change) or a ratio measure (observed belief change ÷ predicted belief change). For the former measure, values closer to zero represent more normative belief updating, whilst for the latter measure, values closer to one represent more normative belief updating. Both measures are, however, susceptible to artifacts stemming from the bounded nature of the probability scale given the potential for response noise (see Section 3.2 for more details).

4.2.4.1. Difference measure. The main effect of Event Type was not significant, $F < 1$, nor was the main effect of Desirability, $F(1,16) = 1.82$, $p = .20$, $\eta_p^2 = .10$. The interaction between the two factors was significant, $F(1,16) = 25.00$, $p < .001$, $\eta_p^2 = .61$. For negative events, participants were significantly less conservative in belief updating in response to desirable ($M = 1.72$ percentage points difference, $SD = 1.95$) than undesirable ($M = 7.62$ percentage points difference, $SD = 5.96$) information, $t(16) = 4.37$, $p < .001$, $d = 1.12$. With regard to positive events, the pattern reversed, as participants were significantly less conservative in their updating in response to undesirable ($M = 3.04$ percentage points difference, $SD = 2.47$) than desirable ($M = 7.07$ percentage points difference, $SD = 4.55$) information, $t(16) = 3.83$, $p = .001$, $d = 1.10$.

4.2.4.2. Ratio measure. Care must be taken when aggregating across multiple trials with the ratio measure as the direction by which a participant differs from the normative change score will alter the weight it is given if the trials are averaged by taking the mean. Should, for example, a participant (for any reason, including keyboard error) update their belief ten times more than predicted, their score on such a trial will be 10, whilst it will be 0.1 if they update their belief ten times less than predicted. Subsequently, the former error will be overestimated if ratios are aggregated across trials by taking the mean. Consequently, for the ratio scores we calculated the median score across trials for each participant and analyzed the data with separate non-parametric Wilcoxon tests for negative and positive events.[11] For negative events, there was a significant effect of Desirability ($Z = 2.59$, $p = .010$), with less conservative updating in relation to desirable information (Median = 0.44; IQR = 0.48) than undesirable information (Median = 0.00; IQR = 0.53). For positive events, there was a significant effect of Desirability ($Z = 2.69$, $p = .007$), with significantly less conservative updating in relation to undesirable information (Median = 0.68; IQR = 0.37) than desirable information (Median = 0.16; IQR = 0.37).

5. Experiment 3A

Experiment 2 both replicated the main finding of Experiment 1 (no evidence for a systematically 'optimistic' pattern of updating when both negative and positive events are considered) and enabled

[11] Note that some data points are lost in the ratio analysis, where the predicted change was zero (for which the ratio would be infinite).

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

examination of the impact of misclassification. Unlike the experiments reported in Sharot and colleagues' studies (Chowdhury et al., 2014; Garrett & Sharot, 2014; Garrett et al., 2014; Korn et al., 2014; Sharot, Guitart-Masip, et al., 2012; Sharot, Kanai, et al., 2012; Sharot et al., 2011), however, Experiments 1 and 2 experimentally manipulated the degree to which participants were inaccurate in their risk estimates by providing base rates derived from participants' initial estimates of personal risk (SE1). The consistency of the results using the personal risk classification scheme for negative events across Experiments 1 and 2, and their consistency with the results reported by Sharot et al. suggests that the use of such derived probabilities did not influence the pattern of results. It is, nevertheless, still possible that this experimental design may have either over- or underestimated the degree to which using the personal risk classification scheme affects the degree of differential updating obtained using the update method. For completeness therefore, as well as to test for asymmetric updating using the normatively appropriate base rate classification scheme on more ecologically valid degrees of accuracy in probability estimation, we used the same conceptual design as Experiment 2 but used externally sourced probabilities, as in Sharot and colleagues' research.

5.1. Method

5.1.1. Participants

Ninety-five UCL psychology undergraduates[12] (76 female; aged 17–21 [median = 19]) participated in Experiment 3A as part of a course requirement. Participants completed the study in two groups in departmental computer laboratories.

5.1.2. Stimuli

Fifty-six (37 negative and 19 positive; Appendix F) life events were presented to participants in one of two random orders. The 37 negative events and associated probabilities were provided by Sharot and Korn and largely overlapped with those used in Sharot et al. (2011).[13] The only changes made to the events were to add the word 'clinical' to obesity (so as to reduce noise in interpretations of obesity) and to present the three different types of cancer separately. The event 'cancer' was also included, with an objective base rate of 40% in England and Wales (Office for National Statistics, 2000).

The positive life events were taken from a previous study (Harris, 2009), and were based on those events used in Weinstein (1980). With the exception of 'graduate with a first', objective statistics were not known for these events, but were taken from participants' estimates in a previous study (Harris, 2009). From the recognition that such estimates are likely to be regressive – that is, less extreme (closer to 50%) than the true statistic (see e.g., Hertwig, Pachur, & Kurzenhäuser, 2005; Moore & Small, 2008), these mean values were then transformed by the equation $\frac{x-.15}{.7}$ to obtain the estimates for the true base rate for these events.[14] Six of these events (marry a millionaire, have a starting salary greater than £40,000, receive nationwide recognition within a profession, have an achievement recognized in the national press, visit the Amazonian rainforest, have one's work recognized with an award) were too rare for this formula to be applied. An estimate that was less regressive (i.e., closer to zero) than the value obtained in Harris' (2009) data was estimated for these events. The statistic for 'graduate with a first' (i.e. the highest possible undergraduate degree classification in the U.K.) was taken as the frequency of first class degrees in the previous year's graduating class (the statistics provided are shown in Appendix F). Participants were informed that the statistics were "from a number of sources and are as reliable

[12] Experiment 3 was conceived independently (by AJLH) of Experiments 1 and 2 (PS & GB). Experiments 1 and 2 used a sample of participants comparable in size to those in Sharot et al. (2011, $N = 19$; Sharot, Guitart-Masip, et al., 2012, $N = 19$ & 21; Sharot, Kanai, et al., 2012, $N = 10$ in each of three experimental groups). Experiment 3 used fewer events than in Sharot et al. (2011) and Experiments 1 and 2, and consequently used a greater sample size.

[13] Experiment 3 used externally sourced probabilities that Korn had kindly provided AJLH for a previous study. Sharot subsequently provided the probabilities used in the Sharot et al. (2011) study. Upon comparison, there was a large degree of overlap (probabilities did not significantly differ, $t(32) = 0.29$, $p = .77$) but not 100% correspondence.

[14] Harris and Hahn (2011) simulated regressive estimates using the formula, $y = 0.7x + 0.15$, which results in estimates (y) that are more regressive than objective probabilities (x), but which equal the objective probability at 0.5. They cited evidence (Clutterbuck, 2008, as cited in Harris & Hahn, 2011) to suggest that this degree of regression was psychologically plausible.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

as they can be". Our base rate estimates for positive events (with the exception of 'graduate with a first') were not therefore taken from external frequency data. It is, however, very difficult to see where one could source such data (at least without an extensive social science survey). Whilst studies have investigated the objective accuracy of people's estimates of the real-world frequency of negative life events (e.g., Christensen-Szalanski, Beck, Christensen-Szalanski, & Koepsell, 1983; Hertwig et al., 2005; Lichtenstein, Slovic, Fischhoff, Layman, & Combs, 1978), we are unaware of any that have investigated this question for positive events. This likely reflects the fact that base rate statistics for positive events are not readily available. We therefore used the procedure above to estimate ostensibly sensible and reliable estimates for the base rates of the positive events.

5.1.3. Procedure

For each event, participants were asked to provide an estimate (as a percentage between 0% and 100%; see Appendix B, Table B3, for mean estimates at each stage) of how likely they thought the event was to occur to them (SE1) and to the average first year UCL psychology student of their age and sex (BR1). In contrast to most previous studies, participants were not constrained to report probabilities in a fixed range only. Rather, participants were free to report a probability anywhere between 0% and 100%, ensuring that participants could report their true beliefs. The order of questions, both the events and whether participants first estimated their own risk or the average person's risk, was counterbalanced across participants. The next screen informed participants of the actual probability that the average first year psychology student would experience that event. To ensure that they processed the information, the following screen asked participants to report the actual probability they had just been given. If the participant entered an incorrect percentage, they were informed that they were wrong, presented with the correct percentage, and required to input that information. The study continued, with one positive event included after every two negative events, until the participant had estimated all events. After rating all 56 events, participants were required to provide a second estimate of their personal likelihood (SE2) and a second base rate estimate (BR2) of experiencing that event. Participants were incentivized to pay attention through informing them that one event would be drawn at random, and one individual whose second base rate estimate was closest to the true value would receive £40. Following the study, participants were debriefed, and informed of the perceived reliability of the sources of the base rate information. A week later, the 'raffle' for the £40 was played in front of all participants, according to the rules above, and the winner was paid in cash.

5.2. Results and discussion

Due to the free nature of participants' responses, it was possible for them to provide percentages that were either less than 0 or greater than 100. There were 6 instances of responses over 100, presumably due to input error. These cases were removed before further analysis. We also excluded trials which were more than 3 interquartile ranges from the mean value from analyses. These exclusions reduced the total number of trials across the four different conditions by approximately 2.5% (the precise number differs across the different classification schemes).

Once again, we first compared the results of the two classification schemes with a 3-way ANOVA, finding a 3-way interaction between Classification Scheme, Event type and Desirability, $F(1,93) = 35.72$, $p < .001$, $\eta_p^2 = .28$. The Event type $\times$ Desirability interaction effect was significantly smaller under the base-rate classification scheme than the personal risk classification scheme, replicating the result observed in Experiment 2. Furthermore, per participant, an average of 6/37 negative events and 3/19 positive events were classified differently under the two classification schemes (i.e., misclassified under the personal risk classification scheme).

5.2.1. Analysis: personal risk classification scheme

Mean updates using the personal risk classification scheme revealed the same asymmetry as observed in Experiments 1 and 2 (Fig. 9A). Average updates were entered into a 2 (Event type: positive, negative) $\times$ 2 (Desirability: desirable, undesirable information) repeated measures ANOVA. The main effect of Desirability was significant, $F(1,93) = 4.45$, $p = .038$, $\eta_p^2 = .05$, but participants updated

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

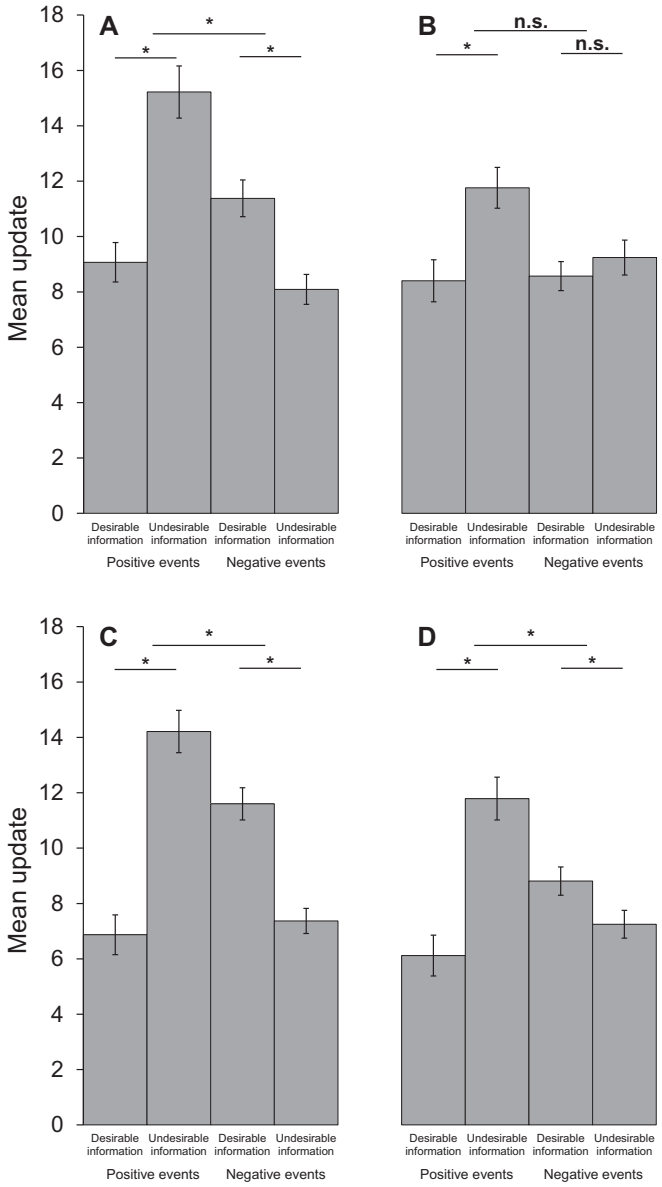

Fig. 9. Mean updates by Event type and Desirability – Experiment 3A. (A) Personal risk classification scheme. (B) Base rate classification scheme. Experiment 3B (C) personal risk classification scheme. (D) Base rate classification scheme. Error bars indicate ±1 standard error of the mean. $^{*}p < .05$, two-tailed.

more in response to *undesirable* information than to desirable information. The main effect of Event type was also significant, with participants updating their estimates about positive events more than negative events, $F(1,93) = 13.34$, $p < .001$, $\eta_p^2 = .13$.

Once again, there was a flip in asymmetric updating across negative and positive events: the significant interaction between Event type and Desirability observed in Experiments 1 and 2 was replicated in Experiment 3A, $F(1,93) = 32.15$, $p < .001$, $\eta_p^2 = .26$. When judging negative events, participants

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

updated more in response to desirable information than undesirable information, $t(93) = 3.54$, $p = .001$, $d = 0.56$), but updated significantly more in response to undesirable information than to desirable information when judging positive events, $t(94) = 5.25$, $p < .001$, $d = 0.75$. As in Experiments 1 and 2, these three critical results remain significant when including a covariate controlling for the initial difference between SE1 and the base rate.

5.2.2. Analysis: base rate classification scheme

The pattern of differential updating is reduced when trials are classified using the normatively appropriate base rate classification scheme (Fig. 9B). Average updates were entered into a 2 (Event type) × 2 (Desirability) repeated measures ANOVA. The main effect of Desirability was significant, $F(1,94) = 13.33$, $p < .001$, $\eta_p^2 = .125$ (updates were greater for *undesirable* information), the main effect of Event type approached significance, $F(1,94) = 3.80$, $p = .054$, $\eta_p^2 = .04$, and there was a trend toward an interaction between Event type and Desirability, $F(1,94) = 2.96$, $p = .089$, $\eta_p^2 = .03$. When estimating the likelihood of negative events, participants updated equally in response to desirable and undesirable information, $t(94) = 0.77$, $p = .44$, $d = 0.12$. When estimating the likelihood of positive events, participants updated significantly more in response to undesirable than desirable information, $t(94) = 3.27$, $p < .001$, $d = 0.46$.

Thus, when trials were classified according to the base rate classification scheme, no evidence of unrealistic optimism was found, although seeming 'pessimism' was observed when judging positive events. To control for the difference in initial base rate errors, the covariate coding for base rate estimation error was again included in the analysis. After inclusion of this covariate, the Event type × Desirability interaction was non-significant ($F < 1$), but the main effect of Desirability remained significant, $F(1,93) = 8.08$, $p < .01$, $\eta_p^2 = .08$.

5.2.3. Analysis: comparisons with rational Bayesian predictions

5.2.3.1. Difference measure. There was a main effect of Event Type, $F(1,94) = 14.89$, $p < .001$, $\eta_p^2 = .14$, such that participants' belief updating was less conservative for positive events ($M = 4.2$ percentage points difference, $SE = 0.6$) than negative events ($M = 6.1$ percentage points difference, $SE = 0.4$). There was no main effect of Desirability, $F(1,94) = 2.75$, $p = .10$, nor was there an Event Type × Desirability interaction ($F < 1$). These results were the same when positive and negative events were analyzed separately (all $ps > .25$).

5.2.3.2. Ratio measure. Separate Wilcoxon related-samples signed rank tests were carried out for negative and positive events. For negative events, there was no effect of Desirability ($Z = 0.27$, $p = .79$), with a trend for less conservative updating in relation to undesirable information (Median = 0.55; IQR = 0.64) than desirable information (Median = 0.48; IQR = 0.77). For positive events, there was again no effect of Desirability ($Z = 1.26$, $p = .21$), with the same trend of less conservative updating in relation to undesirable information (Median = 0.78; IQR = 0.59) than desirable information (Median = 0.56; IQR = 1.00).

5.3. Experiment 3A summary

Experiment 3A included events with externally sourced actual probabilities and observed no evidence for optimistically biased belief updating. Previously, we noted one potential limitation of Experiment 3A, which was that (with the exception of 'graduate with a first') our base rate estimates for positive events were not based on objective frequency data. The consistency of the pattern of results observed across the experiments reported here and in previous studies using the update method, using different methods of generating base rate data for positive events, suggests that this did not affect our findings.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

6. Experiment 3B

6.1. Method

Experiment 3B was a direct replication of Experiment 3A, undertaken a year later. The only difference was the precise demographics of the sample, who again were UCL undergraduates participating as part of a course requirement. 112 participants (91 female; aged 17–22 [median = 19]) were in Experiment 3B.

6.2. Results and discussion

As in Experiment 3A, responses greater than 100 ($n = 4$) were removed before further analysis. We also excluded trials which were more than 3 interquartile ranges from the mean value for that experimental cell from analyses. These exclusions (plus those where a trial could be classified as neither positive nor negative) reduced the total number of trials across the four different conditions by approximately 2.5%.

We first compared the results of the two classification schemes with a 3-way ANOVA, finding a 3-way interaction between Classification Scheme, Event type and Desirability, $F(1,110) = 20.68$, $p < .001$, $\eta_p^2 = .16$. The Event type × Desirability interaction effect was significantly smaller under the base-rate classification scheme than the personal risk classification scheme, replicating the result observed in Experiments 2 and 3A. With respect to misclassification (Section 3.1), an average of 5/37 negative events and 3/19 positive events per participant were classified differently under the two classification schemes.

6.2.1. Analysis: personal risk classification scheme

Mean updates using the personal risk classification scheme revealed the same asymmetry as observed in Experiments 1, 2 and 3A (Fig. 9C). Average updates were entered into a 2 (Event type: positive, negative) × 2 (Desirability: desirable, undesirable information) repeated measures ANOVA. The main effect of Desirability was significant, $F(1,111) = 6.57$, $p = .012$, $\eta_p^2 = .06$, but, as in Experiment 3A, participants updated more in response to *undesirable* information than to desirable information. The main effect of Event type approached significance, with participants trending to update their estimates about positive events more than negative events, $F(1,111) = 3.35$, $p = .07$, $\eta_p^2 = .03$.

Once again, there is a flip in asymmetric updating across negative and positive events: the significant interaction between Event type and Desirability observed in the previous experiments was replicated in Experiment 3B, $F(1,111) = 62.73$, $p < .001$, $\eta_p^2 = .36$. When judging negative events, participants updated more in response to desirable information than undesirable information, $t(111) = 5.12$, $p < .001$, $d = 0.77$, but updated significantly more in response to undesirable information than to desirable information when judging positive events, $t(111) = 6.93$, $p < .001$, $d = 0.93$. As in Experiments 1, 2 and 3A, these three critical results remain significant when including a covariate controlling for the initial difference between SE1 and the base rate.

6.2.2. Analysis: base rate classification scheme

The pattern of differential updating is reduced when trials are classified using the normatively appropriate base rate classification scheme (Fig. 9D). Average updates were entered into a 2 (Event type) × 2 (Desirability) repeated measures ANOVA. The main effect of Desirability was significant, $F(1,110) = 12.85$, $p = .001$, $\eta_p^2 = .11$ (again updates were greater for *undesirable* information), as was the Event type × Desirability interaction, $F(1,110) = 28.28$, $p < .01$, $\eta_p^2 = .21$. When estimating the likelihood of negative events, participants updated more in response to desirable than undesirable information, $t(111) = 2.04$, $p = .044$, $d = 0.29$. When estimating the likelihood of positive events, participants updated more in response to undesirable than desirable information, $t(110) = 5.58$, $p < .001$, $d = 0.71$.

As in previous experiments, the covariate coding for the difference in initial base rate estimation error was again included in the analysis. In this experiment, inclusion of this covariate did not alter the pattern of results in the overall ANCOVA, with the corrected pattern of means still in the direction of pessimism in the overall ANCOVA (undesirable: $M = 9.56$; desirable: $M = 7.40$). Critically, the only

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

result whose significance changed was that the effect of Desirability was no longer significant for negative events, $F(1,110) = 3.55$, $p = .06$.

6.2.3. Analysis: a comparison with rational Bayesian predictions

6.2.3.1. Difference measure. There was a main effect of Event type, $F(1,110) = 7.90$, $p = .006$, $\eta_p^2 = .07$, such that participants' belief updating was less conservative for positive events ($M = 5.2$ percentage points difference, $SE = 0.6$) than negative events ($M = 6.7$ percentage points difference, $SE = 0.4$). The effect of Desirability did not reach the conventional level of significance, $F(1,110) = 3.72$, $p = .056$, $\eta_p^2 = .03$, but the trend was for participants to be less conservative in their updating in response to desirable ($M = 5.4$ percentage points difference, $SE = 0.5$) than undesirable ($M = 6.5$ percentage points difference, $SE = 0.5$) information. There was no Event type $\times$ Desirability interaction ($F < 1$). There was no significant effect of Desirability when negative and positive events were analyzed separately, although, for negative events, the less conservative updating in response to desirable information approached significance, $t(111) = 1.95$, $p = .054$, $d = 0.22$, although note the potential for an inflated Type I error rate with such an analysis in the absence of significant results in the overall ANOVA.

6.2.3.2. Ratio measure. Separate Wilcoxon related-samples signed rank tests were carried out for negative and positive events. For negative events, there was no effect of Desirability ($Z = 1.30$, $p = .19$), with a trend for less conservative updating in relation to desirable information (Median = 0.60; IQR = 0.51) than undesirable information (Median = 0.42; IQR = 0.72). For positive events, there was a significant effect of Desirability ($Z = 2.93$, $p = .003$), but the effect was for less conservative updating in relation to undesirable information (Median = 0.77; IQR = 0.73) than desirable information (Median = 0.28; IQR = 0.88).

7. Experiment 4

The results of all four experiments have been broadly consistent (see Table 3). Whilst Experiments 3A and 3B used externally sourced probabilities, Experiments 1 and 2 derived base rates from initial self estimates. Experiment 4 derived the base rates from initial *base rate* estimates and also attempted to match the subjective frequency of the positive and negative events as much as possible.[15]

7.1. Method

7.1.1. Participants

Thirty-two healthy participants (28 females; aged 18–27 [median = 19]) were recruited via the University of Surrey participant database. All gave informed consent and were paid for their participation.

7.1.2. Stimuli

Forty short descriptions of life events (Appendix G), predominantly taken from Experiment 3, were presented in a random order. The stimulus set was, however, altered so that half of the events were positive and half negative, whilst better equating the base rates of positive and negative events. Specifically, we used the 19 positive events from Experiment 3 (for which we had base rate estimates – see Section 5.1.2). We then added one new positive event, for which there was frequency information for the reference class of intended participants at the University of Surrey (i.e., "Professional or managerial job after graduating" (45%); note we also had frequency information for the event "graduating with a first" (23%), both from unistats.com). This resulted in 20 positive events. A research assistant blind to the experimental hypotheses, with no knowledge of research on optimism bias, was presented with these positive events (and their corresponding base rates – from estimates where necessary, as in Section 5.1.2) and the complete list of negative events used across all experiments (and their corresponding base rates). The research assistant removed events from the list of negative

[15] We thank an anonymous reviewer for suggesting this experiment.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

events in an iterative random manner, until the mean base rates (positive events: 22.75; negative events: 22.85) and standard deviations (23.00; 23.40) were as closely matched as possible ($p = .99$).

7.1.3. Procedure

The procedure followed that of Experiment 2, except that the order in which the estimates were provided was counterbalanced across participants (see Fig. 1B; see Appendix B, Table B5, for mean estimates at each stage). Provided base rates were calculated as ±17–29% of the initial base rate estimate (base rates were capped between 1% and 99% as participants were informed that this was the range of possible probabilities). The funneled debriefing, as implemented in Experiment 1 and 2, revealed that no participants suspected that event base rates were derived from their estimated base rates.

7.2. Results

7.2.1. Comparison of the two classification schemes

A 3-way ANOVA across the two classification schemes demonstrated a trend level 3-way interaction between Classification Scheme, Event type and Desirability, $F(1,28) = 3.40$, $p = .083$, $\eta_p^2 = .11$.

Despite the above interaction failing to attain statistical significance (for the first time), a direct test for misclassification in the data, again revealed significant evidence for this. As in Experiment 2, we entered number of trials into a Classification Scheme × Event type × Desirability repeated measures ANOVA. The three-way interaction between the factors was found to be significant, $F(1,31) = 12.15$, $p = .001$, $\eta_p^2 = .28$, suggesting that a significant proportion of datapoints are misclassified on the personal risk classification scheme (see Section 3.1). Indeed, per participant, an average of 8/20 negative events and 5/20 positive events were classified differently under the two classification schemes.

7.2.2. Analysis: personal risk classification scheme

Data from three participants were omitted from analysis as there were no trials in one or more cells of the design.

Mean updates using the personal risk classification scheme revealed the same asymmetry as observed in the previous experiments (Fig. 10A). Average updates were entered into a 2 (Event type) × 2 (Desirability) repeated measures ANOVA. Neither the main effect of Event type, $F(1,28) = 3.67$, $p = .066$, $\eta_p^2 = .12$, nor that of Desirability, $F(1,28) = 0.39$, $p = .54$, $\eta_p^2 = .01$, was significant.

Once again, there is a flip in asymmetric updating across negative and positive events: the interaction between the factors was significant $F(1,28) = 12.80$, $p = .001$, $\eta_p^2 = .31$. This interaction reflected (non-significant) greater updating in response to desirable than undesirable information when estimating the probability of negative events, $t(28) = 1.87$, $p = .072$, $d = 0.47$, but significantly greater updating in response to undesirable than desirable information when estimating the likelihood of positive events, $t(28) = 2.14$, $p = .041$, $d = 0.37$. Analyses including covariates coding for differences in SE1s, and difference between SE1 and 'actual' base rate, across positive and negative events were conducted and produced the same patterns of significance, as did analyses including only uncapped trials.

7.2.3. Analysis: base rate classification scheme

Although the 3-way interaction above failed to attain statistical significance, the pattern of updating once again appeared somewhat different when trials were classified according to the Base rate classification scheme (Fig. 10B). Average updates were entered into a 2 (Event type: positive, negative) × 2 (Desirability: desirable, undesirable information) repeated measures ANOVA. The main effect of Event type, $F(1,31) = 0.10$, $p = .92$, $\eta_p^2 < .01$ was not significant. However, the main effect of Desirability reached significance, $F(1,31) = 5.53$, $p = .025$, $\eta_p^2 = .15$. The interaction between the factors was not significant in this experiment, $F(1,31) = .58$, $p = .45$, $\eta_p^2 = .02$. The main effect of Desirability arose from participants updating more in response to *undesirable* than desirable information when estimating the likelihood of positive events ($t(31) = 2.24$, $p = .033$, $d = 0.58$), although the simple effect did not reach significance for negative events ($t(31) = 1.09$, $p = .29$, $d = 0.28$). A general optimistic bias clearly cannot account for these results. In fact, if anything, there is evidence of a 'pessimism bias'.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

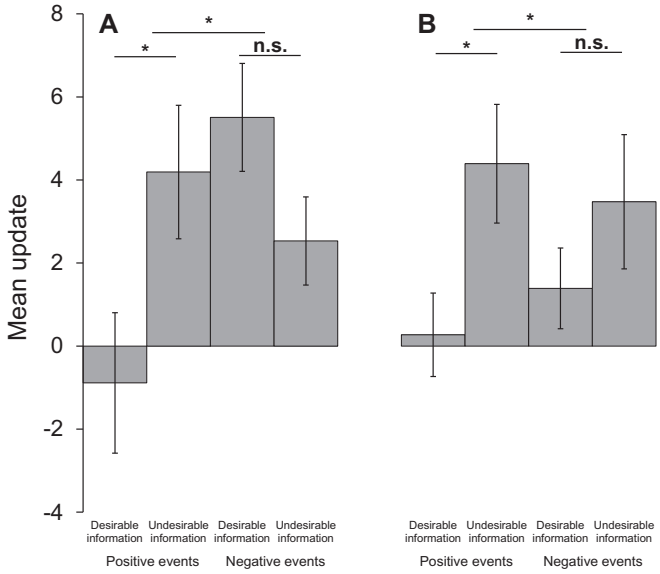

Fig. 10. Mean updates by Event type and Desirability – Experiment 4. (A) Personal risk classification scheme. (B) Base rate classification scheme. No difference in updating was observed under the base rate classification scheme after the degree of base rate estimation error was accounted for. Error bars indicate ±1 standard error of the mean. $^{*}p < .05$, two-tailed.

In an ANCOVA controlling for initial errors in base rate estimates, the main effect of Event Type was not significant ($F < 1$), but the main effect of desirability remained significant, $F(1,30) = 5.34$, $p = .028$, $\eta_p^2 = .15$. The interaction between the two factors was not ($F < 1$). This seeming overall 'pessimism' was not, however, significant for either negative ($F < 1$) or positive events, $F(1,30) = 2.78$, $p = .11$, $\eta_p^2 = .09$, individually.

7.2.4. Analysis: a comparison with rational Bayesian predictions

7.2.4.1. Difference measure. There were no significant main effects of Event type ($F < 1$), or Desirability, $F(1,31) = 3.74$, $p = .076$, $\eta_p^2 = .09$, and the Event type × Desirability interaction was not significant ($F < 1$). There was, however, a trend for participants to be less conservative in their updating in response to undesirable ($M = 2.45$ percentage points difference, $SE = 0.91$) than desirable ($M = 4.32$ percentage points difference, $SE = 0.65$) information.

7.2.4.2. Ratio measure. Separate Wilcoxon related-samples signed rank tests were carried out for negative and positive events. For negative events, there was no effect of Desirability ($Z = 1.64$, $p = .098$), with a trend for less conservative updating in relation to undesirable information (Median = 0.32; IQR = 1.24) than desirable information (Median = 0.00; IQR = 1.26). For positive events, there was no effect of Desirability ($Z = 1.02$, $p = .31$), but the trend was for less conservative updating in relation to undesirable information (Median = 0.39; IQR = 1.11) than desirable information (Median = 0.00; IQR = 0.97).

7.3. Robustness and anticipated objections

Across our experiments we find consistent evidence against 'optimism' in both negative and positive events, with significant interactions by event type (positive vs. negative) in every study. This is consistent with the statistical artifacts that plague the update method, and it is not consistent with a genuine optimism account. In a sense, no experimental data are needed to demonstrate that the update method is flawed – for this, the simulations of Section 3 suffice. Nevertheless, the experiments

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

demonstrate the difficulties further. The experiments are also informative, we believe, because they vary in procedural details, sample size, and, to a certain extent, the events included.

Of these factors, the one that *should* have a large effect on what is observed are the events. This too should be clear from Section 3. Future life events will inevitably vary in base rate and in the amount of individual diagnostic information that participants have at their disposal. It is worth stressing this point once more, because it cuts off a potential line of fruitless debate: namely, the idea that optimism only fails to emerge for positive events in our experiments because we have somehow not chosen quite the right events, or because positive events are genuinely different in some other way. "Unrealistic optimism about future life events" (e.g., Weinstein, 1980) should be about exactly that: future life events, not some future life events on some occasions in some people (Hahn & Harris, 2014). Asymmetric updating was introduced by Sharot et al. (2011) as a way of explaining "How unrealistic optimism is maintained in the face of reality" and the same generality should apply here too. Asymmetric updating simply fails as an explanatory mechanism for generating systematic optimism if it leads to optimism for some events, but to pessimism for others.

At the same time, it is important to remember that the statistical artifacts will necessarily vary in expression as a function of events: in other words, it will be possible to find some set of events that will generate the desired 'seeming optimism' both in negative and positive events. Research could easily converge on 'just the right' items, which would give rise to consistent (artifactual) optimism, if this were a worthwhile goal. The experimental data to date, coupled with the insight provided by the simulations, would help construct such a set. Fig. 11 shows the distribution of different events by base rate in Sharot et al.'s seminal (2011) study. As can be seen, they are predominantly rare events, which will give rise to likelihood ratios below 1 in the majority of participants. And these are the event characteristics on which we based our simulations.

It must be stressed, however, that while event frequency is important, it does not have the same direct relationship with artifactual optimism as it does for the classic comparison method of Weinstein and colleagues (see Section 3, Harris & Hahn, 2011; Weinstein, 1980). It matters whether events are above or below 50% (i.e., frequent or rare), but there is no simple relationship between the size of the artifact introduced by Sharot et al.'s normatively inappropriate procedure and event rarity, because the magnitude of the base rate error and the exact likelihood ratios (participants' individual diagnostic knowledge) matter as well (as the discussion of Table 2 in Section 3 demonstrates).

What should (artifactually) generate reliable 'optimism' for positive events, however, is a set of positive stimuli that was exactly the complement of Sharot et al.'s (2011) set with respect to base rates, while exactly matched for diagnostic knowledge. In fact, Sharot et al. elicited responses to such

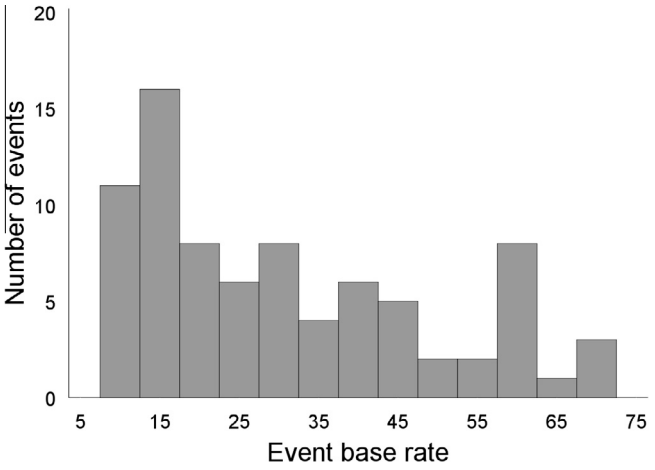

Fig. 11. Distribution of event frequencies in Sharot et al. (2011).

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

a set by asking another set of participants about their chance of *not experiencing* the relevant events. And this is exactly what Sharot et al. found (see Section 2).

In summary, the very nature of the artifacts that plague the update method mean that, given the right set of events, everything and anything could empirically be found, even in entirely unbiased agents. As a result, a search for 'better' events is not the way to go.

If anything is going to save the update method methodologically, it will have to be new measures that are calculated from the data. We stressed in Section 3 the limitations of all of the measures we could think of, including the method of comparing predicted and observed posteriors. Other methods might yet be found, reviving the paradigm. It would be a great boost to optimism research if this were possible, but we are presently unaware of an approach that would fit the bill. In this context, it is also worth discussing the other dependent variable reported in a number of papers using the update method (including the original, Sharot et al., 2011). This is the 'learning score' (Moutsiana et al., 2013). As indicated in Section 3, the learning score is "the correlation between estimation errors and subsequent updates across trials" (Moutsiana et al., 2013, p. 16397), where estimation error is 'SE1 – actual base rate' in the current terminology. The simulation presented in Section 3.2, and the data in Table 2, demonstrate that asymmetric learning scores for desirable versus undesirable trials also readily arise from unbiased, Bayesian agents. Thus, the learning score is not immune from the problems identified with the update method. In our experiments, however, we only report update scores. The reason for this is straightforward. We are concerned with 'optimistic' updating. A difference between the 'learning scores' for desirable and undesirable trials does not reflect optimistic updating, as can be demonstrated through a simple example.

The learning score is a correlation. Any correlation is insensitive to absolute value. It is only the relationship between the values that is of consequence. Assuming that they are not 'overshooting', an individual who updates three times as much in response to desirable information than undesirable information should presumably be characterized as optimistically updating (assuming, here, that the logic for the rest of the method were completely unflawed). This result will not necessarily be observed in a correlation between initial error and amount of update. If this optimistic updater always updates 1 unit for every 10 units of initial error in undesirable trials, but in desirable trials they update 3 units for every 10 units of initial error, the learning score (as a correlation) will not distinguish between these two conditions (both correlations will equal 1).

In other words, the learning score not only shares the problems of update scores, with respect to the inability to control for both base rate error and individual diagnostic knowledge, but it is also conceptually limited with respect to the question of how people respond to positive as opposed to negative error signals. Consequently, the search for appropriate measures will need to take other directions.

8. General discussion

The current paper presented a detailed exploration of the phenomenon of optimistic belief updating in probability estimates through an analysis of the logic of the update method (e.g. Sharot et al., 2011), simulations and experimental data. Experiment 1 replicated the optimistic updating effect with estimates about the occurrence of negative events, but showed seeming *pessimistic* updating with positive events. Following this, we presented a methodological critique of the updating method and demonstrated that seemingly optimistic and pessimistic updating could be observed from a simulated population of rational Bayesian agents. In addition to demonstrating that the original updating task defines new probability information inappropriately (with information about the base rate defined as desirable or undesirable based on its relationship with people's estimates about their *own* personal risk, rather than their perception of the base rate), it was shown that the bounded nature of the probability scale currently prohibits a perfect, principled way of investigating the research question. This is because diagnostic information that people are in possession of will have differential effects at different parts of the probability scale – thus preventing a meaningful comparison between trials on which event likelihoods are underestimated and those on which they are overestimated. Furthermore, these effects differ according to the base rates of the events under consideration, as well as the events

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

themselves (differential amounts of diagnostic information will be available for different events, and for different people about different events). Consequently, in four further experiments we employed a range of measures (methods, analyses, events) to provide the most rigorous test to date of the optimistic belief updating hypothesis. Despite the problems inherent in the update method, should a belief updating pattern consistent with optimism have been observed across a range of different events and analyses, one may have been tempted to conclude that optimistic belief updating exists. As Table 3 demonstrates, however, this was not the pattern of results observed. This, coupled with the simulations of rational Bayesian agents, means that the most parsimonious explanation is that optimistic belief updating is an artifact of flaws in the methodology used, together with the bounded nature of the probability scale.

One thing that has been highlighted by our exploration is the importance of recognizing what the numerical responses obtained from participants represent. In studies that require participants to consider and report their own risk of experiencing real events in the real world, the response from participants is more complex than most responses in psychological experiments. Participants' simple numerical responses may contain a wealth of information (indeed, the normative theory suggests they should), all of which is outside of the control, or even knowledge, of the experimenter. Recognizing this was at the heart of the critique of comparative unrealistic optimism research proffered in Harris and Hahn (2011). The simulations presented in the current paper again highlight the importance of considering these factors when determining what can and what cannot be learnt from such methodologies.

The fact that there is a normative theory of how risk judgments should be made is consequential both for understanding the task that participants face and for analyzing their responses. Sharot et al.'s seminal (2011) study overlooks the fact that unbiased, rational participants should combine base rate information with individual diagnostic knowledge. This not only leads to systematic misclassifications of participant trials, but also affects the ways in which one can control for random differences in initial error. One needs to control for (theoretically uninteresting) random differences in participants' initial error between desirable and undesirable trials in order for differences in subsequent update to be theoretically meaningful. Given that base rate information is the only information provided in these studies, it is initial base rate error that should be controlled for and thus matched. It might seem intuitive that, in the absence of other systematic differences, equal amounts of initial base rate error should result in equal amounts of updating. However, as Section 3.2 showed, once participants have some individual diagnostic knowledge, it is simply impossible to match participants' individual diagnostic information, participants' base rate error and the amount by which they have to update their beliefs. Only two of these three can be simultaneously matched across those receiving desirable and undesirable information. Consequently, if one matches base rate error and diagnostic information, amount of update will necessarily be asymmetric. Of course, as soon as base rate error or diagnostic information differ, the amount of update should differ. This underlines the problems for any attempt to compare the amount of update across trials in which estimates should move in opposite directions in different parts of the probability scale, as is the case when comparing the updates of participants receiving desirable and undesirable information.

In light of these difficulties, it is of critical importance to utilize tests that incorporate a variety of different events. The inclusion of positive events (with similar base rates to the negative events) can act as a simple litmus test to determine whether a test is likely demonstrating true evidence of valence-biased judgment. Of course, it is possible for a genuine result to reveal people to be optimistic for judgments of negative events and pessimistic for positive events (and vice versa). It should, however, serve as a warning to thoroughly investigate the characteristics of the methodology used. Simulations of rational agents being put through the test should yield data that seem indicative of rationality. If rational agents appear biased, one cannot make any claims from the same pattern of data obtained from human participants.

The most important inclusion, therefore, in all the studies presented is the positive events. Confounds associated with the bounded probability scale are reversed for positive events (as higher and lower probabilities become more and less desirable events respectively – rather than less and more, as they are for negative events). Thus, the inclusion of these events is critical to enable any conclusion to be drawn whatsoever about optimistic updating (as a general characteristic of human

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

thought, rather than potentially something solely relating to negative events – on the importance of such generality see, Hahn & Harris, 2014).

8.1. How do people update their beliefs?

In the simulations reported in this paper, we demonstrate that rational Bayesian agents can produce data that seem indicative of optimistic belief updating. By definition, these agents are not optimistic, and it thus follows that the same pattern of results cannot necessarily be taken as evidence of optimism in human participants. This does not, however, necessarily mean that humans *are* rational Bayesian agents, and it certainly does not mean that a process account of human belief updating would resemble Bayes' Theorem. To our minds, a process account of belief updating is not yet sufficiently developed for it to be utilized in developing a test of *optimistic* belief updating. Potential candidates in the literature for process theories of belief updating include: Explanatory coherence (e.g., Thagard, 1989, 2000), averaging processes (e.g., Juslin, Nilsson, & Winman, 2009), associative processes and natural assessments (e.g., Morewedge & Kahneman, 2010), and Bayesian process models (e.g., Sanborn, Griffiths, & Navarro, 2010). Any or none of these might subsequently be endorsed as *the* underlying process of human belief updating.

The fact that we do not know whether or not people update as Bayesian agents, however, is moot with respect to the role of the simulations in the present paper. The central conclusion that has previously been reached from studies using the update method is that human belief updating is (typically) optimistic, in that people update more in response to desirable information than undesirable information (with negative events). This conclusion relies on the assumption that non-optimistic agents would update equally in response to desirable and undesirable information. Bayesian agents are not optimistic. That simulations with Bayesian agents show unequal belief updating in response to desirable and undesirable information thus demonstrates that this critical assumption of the update method does not hold.

Without that critical assumption, however, none of the further empirical conclusions that research within update paradigm has put forward can be maintained. It becomes entirely unclear, for example, what the neural correlates represent, what L-DOPA moderates in this context, or what effects of ageing on updating are (Chowdhury et al., 2014; Garrett et al., 2014; Sharot, Guitart-Masip, et al., 2012; Sharot, Kanai, et al., 2012; Sharot et al., 2007, 2011). At the same time, the simulations make clear that the mere fact that 'optimistic updating' is subject to moderation by TMS (Sharot, Kanai, et al., 2012), L-DOPA (Sharot, Guitart-Masip, et al., 2012), age (Chowdhury et al., 2014), or even the presence of depressive symptoms (Garrett et al., 2014) does not imply that the effect is somehow real. As the sharp boundaries seen in Fig. 3 (Section 3.1) above demonstrated, any change in participants' perception of the underlying probabilistic quantities or the degree of noise with which participants report them, is likely to give rise to seemingly 'qualitative' differences. It remains entirely possible that any or all of these identified moderators do increase or decrease optimism, but this conclusion cannot be made from the extant data using this methodology. All of this serves to illustrate the extent to which any process level analysis can proceed only with a full understanding of the nature of the task at a computational level which draws out the task requirements, the goals of the individual and how these are optimally achieved (see also e.g., Oaksford, 2015).

8.2. Unrealistic optimism about future events?

Finally, our results have a number of implications for the current debate about unrealistic optimism itself. In that debate, there are presently a number of fundamental questions about unrealistic optimism that must be resolved. It is beyond doubt that the standard method of optimism research, the comparative method (Weinstein, 1980), generates data that *seem* indicative of unrealistic optimism for negative events. Regardless of the appropriate interpretation of these data, at a group level participants rate themselves as less likely than the average person to experience adverse future life events. Yet the following three, interrelated, questions are presently without compelling answer: (1) how is that pattern brought about? (2) does it reflect optimism on the part of the participants? and (3) if yes, is that optimism an instance of a general optimistic bias with adaptive properties?

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

The idea that optimism is a cognitive illusion that promotes well-being (e.g., Sharot, 2012; Taylor & Brown, 1988) straightforwardly says 'yes' to (2) and (3), and implies a valence-based mechanism such as motivated reasoning that will reliably generate optimism to underlie (1) (Kunda, 1990; Lench & Bench, 2012; Weinstein, 1984). However, a number of alternative proposals for mechanism exist in the literature, which maintain that comparative optimism may arise from general (non-motivational) constraints on the comparison process. On the egocentrism account (e.g., Chambers et al., 2003), it is the fact that participants will naturally focus on themselves and their own individuating information that gives rise to comparative optimism. On the differential regression account (Moore & Healy, 2008; Moore & Small, 2008), it is the fact that participants have more accurate knowledge about the self than about the average person that gives rise to the discrepancy. On these accounts, participants may be comparatively optimistic (i.e., a 'yes' to (2) above), but may, on other occasions, equally show comparative pessimism (as is indeed observed for positive events of comparable frequency, Chambers et al., 2003; Harris, 2009; Kruger & Burrus, 2004; Moore & Small, 2008), thus rejecting a general optimism bias (saying 'no' to 3). The statistical artifact hypothesis (Harris & Hahn, 2011), finally, suggests that the seeming comparative optimism seen in traditional optimism studies merely reflects artifactual, method-based distortions of what are actually unbiased responses.

The present results are relevant to each of these three questions and these different accounts. A demonstration of (seeming) robust pessimism for positive events (as observed in four out of our five studies under the personal risk classification) is incompatible with the idea that there is a general, self-protective optimism bias and is also incompatible with any kind of motivated reasoning account. Ironically then, our critique of the methodology, which questions the meaning of this 'pessimism', is the thing that keeps these accounts alive. Only these methodological limitations prevent a firm conclusion that participants' belief updating is pessimistic for positive events, which would have ruled out such accounts. The lack of definitive evidence for pessimistic updating allows, at least in principle, that a general optimistic bias may be generated by means other than differential updating. Of course, it continues to be unexplained on such an account, however, why the comparative method (e.g., Weinstein, 1980) yields pessimism for rare positive events (Chambers et al., 2003; Harris, 2009; Kruger & Burrus, 2004; Moore & Small, 2008).

8.2.1. Implications for automatic optimism

The failure to find any evidence for optimistic belief updating does, however, present a serious challenge to one particular motivational account: Lench and colleagues' 'automatic optimism' (Lench, 2009; Lench & Bench, 2012; Lench & Ditto, 2008). This account assumes that optimism arises from automatic affective responding. Specifically, people make use of an 'optimism heuristic'. People "simply decide that events that elicit positive affective reactions are likely to occur and events that elicit negative affective reactions are unlikely to occur" (Lench & Bench, 2012, p. 351) as a default reaction when they experience approach or avoidance reactions to future events.

In particular, this means that participants will interpret base rates optimistically (Lench & Ditto, 2008). In support, Lench and Ditto (2008) provided participants with matched positive and negative future life events, for which participants were given equal base rates. Participants then rated their own chance of experiencing that event. In a direct comparison, the mean estimates for negative events were lower than for the matched positive events, suggesting optimism.

In principle, this is an interesting test, because it does not rely on the comparative method. But one cannot test for optimism in this way unless – as an experimenter – one's knowledge of the base rate is entirely accurate, and these true base rates are exactly equal across conditions. This is patently not the case in Lench and Ditto's (2008) design, because they used negation to generate corresponding negative ('will get cancer') and positive ('will not get cancer') events.

Complementary events can be equiprobable only if their base rate is exactly 50%. However, this is not the case for events such as 'getting cancer', 'owning one's own home', 'at some point being unemployed', or 'developing asthma' as used by Lench and Ditto (2008). That participants are told equivalent base rates is immaterial here, because the distribution of participants' individual diagnostic knowledge will be governed by the true base rate (precisely because that knowledge is diagnostic). For example, cancer has a life-time prevalence of about 40%, so most people will genuinely not get cancer. By the same token, more people will possess diagnostic knowledge indicating lower risk than

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

there will be people with knowledge indicating greater risk (see also Harris & Hahn, 2011, Section 3 and Appendix D). This means that averages across estimates of individual risk will deviate from the experimenter provided base rate even if participants fully believe that base rate and incorporate it into their own prediction. Once more, it is critically important not to lose sight of what these numbers actually represent for participants in the real world.

8.2.2. Implications for Harris and Hahn's (2011) statistical artifact hypothesis

Needless to say, evidence for a genuine desirability bias in belief updating would have conflicted with the statistical artifact hypothesis concerning comparative optimism. Selectively revising beliefs in response to desirable as opposed to undesirable information would drive down estimates of individual risk for negative events relative to what they should be (and raise those estimates for positive events). On average, over time, this would give rise to genuine comparative optimism. The examination of optimistic belief updating conducted here thus presents one critical test of the statistical artifact hypothesis. The failure to find any evidence of selective updating is in line with that hypothesis. The experiments reported here also provide indirect evidence supporting the statistical artifact account. Harris and Hahn (2011) showed through simulation that the comparative method makes entirely rational agents seem optimistically biased. This renders useless the evidence from comparative tests, because one cannot establish the existence of irrational bias with measures that yield seemingly biased data from unbiased agents. This, however, leaves open the extent to which participants actually are unbiased agents, and whether there is also genuine optimism influencing the outcome of such tests. In order for it to be the case that statistical artifacts are all there is to comparative optimism data, participants must share some of the fundamental characteristics of rational Bayesian agents. While they need in no way be perfect Bayesians, participants would have to be sensitive to base rates, possess individuating knowledge, and make use of both in their estimates of personal risk. The statistical artifacts identified by Harris and Hahn (2011) would exert no effect if participants felt they had no individuating knowledge with which to distinguish themselves from the base rate (the average person's risk). The same is true if they were entirely unaware of base rates, or failed to incorporate them into their own estimates of personal risk as has sometimes been claimed (e.g., Kahneman & Tversky, 1973; but see also Cosmides & Tooby, 1996; Welsh & Navarro, 2012). The data from Experiments 2–4 make clear that this is not the case. People readily incorporate a new base rate into their distinct estimate of individual risk. Though one cannot, at this point, conclude that comparative optimism is solely a statistical artifact, it remains entirely possible that it is.

8.2.3. A broader perspective on unrealistic optimism

Against this conjecture it may seem tempting to argue that support for a genuine optimistic bias may be drawn from other sources that do not involve estimates of risk (Dunning, Heath, & Suls, 2004). Specifically, unrealistic optimism about risks for future events is often linked to an array of self-serving biases, including overconfidence (e.g., Kahneman & Tversky, 1973; Lichtenstein, Fischhoff, & Phillips, 1982) and the better-than-average effect (e.g., Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995; Svenson, 1981). If there is robust evidence of very broad self-enhancement biases, one may be inclined to consider it unlikely that unrealistic optimism does not in fact exist. However, these other biases have also drawn critical attention, with their classification as self-enhancement biases questioned by a significant number of researchers (e.g., on overconfidence: Erev, Wallsten, & Budescu, 1994; Hogarth & Karelaia, 2011; Moore & Healy, 2008; Pfeifer, 1994; Soll, 1996. On the better-than-average effect: Benoit & Dubra, 2011; Galesic, Olsson, & Rieskamp, 2012; Kruger, 1999; Kruger, Windschitl, Burrus, Fessel, & Chambers, 2008; Moore & Healy, 2008; Moore & Small, 2007). It could also be argued that unrealistic optimism is related to general self-serving biases in belief updating, such as a selective failure to incorporate negative information pertaining to one's own attractiveness (Eil & Rao, 2011).

However, even if the preceding biases are found to be robust (and recent studies controlling for extant critiques suggest this may be true for certain types of overconfidence: Benoit, Dubra, & Moore, 2015; Mannes & Moore, 2013; Merkle & Weber, 2011), there is no reason why risk estimates should be optimistic. There might be little harm in overestimating one's own physical attractiveness. It might even be beneficial for "high ability" college students to conservatively update their expectations

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

of future earnings in the light of undesirable information (Wiswall & Zafar, 2015; but see, Tenney, Logg, & Moore, 2015, for the possibility that research might have exaggerated the potential for such a benefit), so as to maintain motivation and subsequently maximize future earnings. One can argue, however, that such optimism is likely to be more harmful when estimating risk. The costs associated with underestimating the probability of a negative event will often be greater than the costs associated with overestimating it, as people will not take the necessary protective behaviors (e.g., Harris, Corner, & Hahn, 2009; Weber, 1994) ("people who play down the seriousness of early symptoms of severe diseases such as skin cancer. . .may literally pay with their lives for their motivated reasoning"; Kunda, 1990, p. 496).

Estimates of the probability of experiencing future life events may be viewed not only in a broader context of potentially self-enhancing biases, but also within the broader context of research on probability judgment. In the same way that there exists a wider literature on self-enhancement, there exists a broader literature examining the effects of desirability on estimates of probability, and the broader picture emerging from that literature seems equally relevant.

Here, research on unrealistic optimism naturally links to a broader swathe of research that is usually brought under the banner of 'wishful thinking'. Across a variety of paradigms investigating likelihood judgments for future events, some researchers have reported evidence for optimism (e.g., Irwin, 1953; Lench, 2009; Pruitt & Hoge, 1965), some have demonstrated that negative events are overestimated (suggesting pessimism) (Bilgin, 2012; Harris et al., 2009; Risen & Gilovich, 2007), some have demonstrated an overestimate of both positive and negative events (Vosgerau, 2010; but see de Molière & Harris, 2016), whilst others have observed no effect of outcome utility, or established that a seeming effect of outcome utility is merely an effect of salience (Bar-Hillel & Budescu, 1995; Bar-Hillel, Budescu, & Amar, 2008). Given the discrepancy of research findings, and alternative interpretations of many 'optimistic' findings to date (e.g., Bar-Hillel et al., 2008; Harris et al., 2009; Windschitl, Smith, Rose, & Krizan, 2010; see also, Krizan & Windschitl, 2007), systematic research into this area identifying clear methods is of paramount importance. At present, however, it seems, to us, premature to conclude that people's judgments of probability are systematically prone to 'wishful thinking'. Viewed from the wider perspective of judgment and decision-making research, the evidence for unrealistic optimism is far from overwhelming.

Although we – on the basis of the divergent findings cited above, Harris and Hahn's (2011) critique, and the present studies – doubt the existence of a general optimistic bias, it is almost certainly the case that there will be optimism in specific cases. Specific groups of people, for example smokers (Arnett, 2000; Weinstein, 1998) and gamblers (Gibson & Sanbonmatsu, 2004), *might* be more optimistic than others. People also might be optimistic about specific things. For example, there has been convincing evidence of 'costly' optimism in sports fans (Babad & Katz, 1991; Massey, Simmons, & Armor, 2011; Simmons & Massey, 2012; but see, Bar-Hillel et al., 2008).

Moreover, beyond the domain of future life events, people might be overoptimistic about specific things because they are, in fact, optimistic in their updating. Evidence for optimistic updating has, for example, been found in people's perceptions of their attractiveness (Eil & Rao, 2011), personal traits (e.g., Korn, Prehn, Park, Walter, & Heekeren, 2012), and their intelligence (Eil & Rao, 2011; Mobius, Niederle, Niehaus, & Rosenblat, 2011). Once more, it is important to bear in mind the fundamental differences between personal risk and things such as attractiveness or personal traits. This matters because Kunda (1990), in her seminal review of motivated reasoning, stressed that people will be more likely to arrive at desired conclusions only if they are able to construct or retrieve evidence in their support which is further helped by properties such as stimulus ambiguity (cf. Ditto & Lopez, 1992; Dunning et al., 1989). Multifaceted concepts such as (unspecific) intelligence and attractiveness can be interpreted in many ways, which is not the case for whether or not a person will contract cancer. As just discussed, wishful thinking has proved difficult to establish in the judgment and decision-making literature. This literature makes use of minimal materials that are aimed at demonstrating a pure desirability bias that does not rest on the fact that biased evidence has been accumulated. The current studies involving the update method share the same fundamental characteristics: participants are provided with a single, simple statistic about the event base rate, that it seems impossible to construe in more than one way. It is thus entirely in keeping with the findings of both the judgment and decision-making literature (discussed above) and the literature on motivated reasoning that no opti-

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

mistic update bias has been found. There may still be genuine unrealistic optimism about specific future life events. This may also turn out to be brought about by updating asymmetries. However, the literature on motivated reasoning itself suggests that such asymmetries (and resultant bias) are likely to be found only in far richer information contexts – contexts that can be interpreted in a variety of different ways.

We conclude, therefore, that there is presently no compelling evidence for a fundamental optimism bias by which people, in general, are optimistic about future life events, in general. Until new, independent evidence for such a bias can be obtained, it seems prudent to revert to an accurate perception of reality as our gold standard for mental health (see also, Colvin & Block, 1994), rather than assuming that optimism is a general human trait that promotes mental wellbeing (e.g., Sharot, 2012; Taylor & Brown, 1988).

8.3. Conclusion

In light of recent critiques, new tests of optimism are of great interest, especially given their application to real-world decision making in the fields of health and finance. A recent test (Chowdhury et al., 2014; Garrett & Sharot, 2014; Garrett et al., 2014; Korn et al., 2014; Kuzmanovic et al., 2015, 2016; Moutsiana et al., 2013; Sharot, Guitart-Masip, et al., 2012; Sharot, Kanai, et al., 2012; Sharot et al., 2011) was shown to be inappropriate in its original formulation. A detailed analysis highlighted further difficulties inherent in assessing the (ir)rationality of belief updating about real-world events in this and other contexts. Five experiments tested for optimistic belief updating through the use of a variety of events (both positive and negative) and via a number of analyses. Overall, no evidence for optimistic updating was observed. The results of Sharot and colleagues thus do not provide additional evidence in favor of a general human optimistic tendency. Furthermore, even if subsequent research does point to the universality of unrealistic optimism, the underlying mechanisms and neurobiology supporting the phenomenon remain unknown.

Acknowledgments

We thank Tali Sharot and Christoph Korn for providing a list of the negative life events and their associated probabilities. A preliminary report of the first simulation was reported in the proceedings of CogSci 2013. Punit Shah acknowledges the support of the Experimental Psychology Society, the Medical Research Council and the Wellcome Trust [099775/Z/12/Z].

The project was conceived by all authors. GB and PS designed, conducted, and analyzed Experiments 1 and 2. AJLH and UH undertook the rational analyses and ran the simulations. AJLH designed, conducted and analyzed Experiments 3A and 3B. CC, GB, and PS designed Experiment 4, conducted by CC and analyzed by CC, GB and PS. All authors drafted and commented on the manuscript.

Appendix A. Life events – Experiment 1

A warm sunny day in winter
Appendicitis
Attending a friend's birthday party
Being blackmailed
Being cheated on by partner
Being convicted of crime
Being fired
Being in unmanageable debt
Being offered a seat on a busy train
Being on the queen's birthday honours list (OBE/CBE/Knighthood)
Being told you are special

(*continued on next page*)

Being told you look attractive/handsome
Bone fracture
Buying a new home
Cancer (of digestive system/lung/prostate/breast/skin)
Vehicle (car/bike) stolen
Card fraud
Chronic high blood pressure
Computer crash with loss of important data
Constant healthy weight for 10 years
Death before 70
Death by infection
Death of a pet animal
Deep vein thrombosis/blot clot in vein
Dementia
Divorce
Domestic burglary
Eating at your favourite restaurant
Falling down stairs
Family visit you at Christmas
Finding money in a coat pocket
Finding something valuable you thought you had lost
Fraud when buying something on the internet
Free trip around the world
Getting a large bonus payment at work
Getting a new pet animal
Getting engaged/married
Getting severely sunburnt
Going to a movie premiere
Having a book/article published
Having a child
Having a stroke
Heart failure
Helped by a stranger when you need it
House vandalized
Inheriting a fortune unexpectedly
Limb amputation
Living healthily past 80
Losing wallet
Marrying someone wealthy
Meeting a member of the royal family
Miss a flight
Mouse/rat in house
Not getting ill all winter
Parkinson's disease
Promotion/new job
Receiving a present
Receiving an unexpected discount on a purchase
Running a profitable business
Running into a long lost friend
Seeing someone famous while on public transport
Severe injury due to accident (traffic or house)
Sexual dysfunction

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

Starting a new relationship
Theft from person
Upgraded to first class on a flight
Theft from vehicle
Victim of bullying at work (nonphysical)
Victim of mugging
Victim of violence by acquaintance
Victim of violence by stranger
Victim of violence with need to go to A&E
Winning a prize in a media (TV/radio/newspaper) competition
Winning a race
Winning a raffle at a fair
Winning a raffle for a convertible sports car
Winning an all-expenses paid holiday
Winning an award in recognition of your work
Witnessing a traumatising accident
Your achievements in newspaper

A.1. Life events used during practice trials

Glaucoma
Winning the lottery

Appendix B. Descriptive statistics for estimates (not updates) in all experiments

Events coded as desirable or undesirable on the personal risk classification scheme. Actual BR is the base rate information provided to the participants by the experimenter. SE1 is initial personal risk estimate, SE2 is the updated estimate of personal risk. BR1 is initial base rate estimate, BR2 is the second base rate estimate. Mean estimates and standard deviations (in parentheses) were, like belief updates, calculated by averaging information about each trial type at the group level (see Tables B1–B5).

Table B1
Experiment 1.

	Positive events			Negative events		
	Undesirable trials	Desirable trials	Overall	Undesirable trials	Desirable trials	Overall
Actual BR	28.39 (7.28)	38.00 (12.70)	33.20 (6.47)	35.30 (14.05)	22.19 (9.87)	28.74 (11.30)
SE1	39.46 (9.87)	31.43 (12.36)	35.44 (7.28)	28.65 (11.93)	30.45 (12.77)	29.55 (11.62)
SE2	33.08 (8.90)	33.79 (11.88)	33.43 (6.41)	30.31 (13.06)	24.55 (11.42)	27.43 (11.67)

Table B2
Experiment 2.

	Positive events			Negative events		
	Undesirable trials	Desirable trials	Overall	Undesirable trials	Desirable trials	Overall
Actual BR	21.17 (7.59)	37.80 (15.07)	29.48 (10.52)	32.10 (15.07)	16.24 (8.55)	24.17 (10.87)
BR1	33.33 (9.88)	28.24 (12.08)	30.79 (10.38)	24.53 (12.32)	26.72 (12.48)	25.62 (11.90)
SE1	28.21 (11.14)	31.53 (12.39)	29.87 (10.74)	26.60 (12.89)	20.66 (12.34)	23.63 (11.17)
SE2	22.72 (10.35)	32.76 (14.21)	27.74 (11.05)	24.62 (13.58)	15.08 (9.40)	19.85 (10.22)

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

Table B3
Experiment 3A.

	Positive events			Negative events		
	Undesirable trials	Desirable trials	Overall	Undesirable trials	Desirable trials	Overall
Actual BR	15.28 (5.25)	31.90 (7.74)	22.25 (1.63)	32.27 (3.94)	15.16 (2.33)	23.22 (0.93)
BR1	33.64 (7.40)	17.90 (5.89)	26.72 (5.37)	14.12 (3.42)	31.74 (5.60)	23.29 (4.93)
SE1	33.78 (10.92)	20.53 (10.06)	27.98 (9.17)	13.27 (5.67)	28.42 (9.31)	21.20 (7.55)
BR2	17.79 (7.01)	31.86 (8.74)	23.83 (4.63)	28.89 (6.53)	20.77 (5.72)	24.50 (4.92)
SE2	22.01 (11.09)	28.94 (10.80)	25.07 (8.48)	22.51 (8.16)	19.87 (8.29)	21.03 (7.32)

Table B4
Experiment 3B.

	Positive events			Negative events		
	Undesirable trials	Desirable trials	Overall	Undesirable trials	Desirable trials	Overall
Actual BR	16.40 (5.09)	30.59 (8.70)	21.98 (1.49)	31.07 (4.89)	15.24 (2.64)	22.72 (1.27)
BR1	35.93 (8.84)	17.53 (8.08)	28.58 (7.31)	13.83 (4.20)	32.97 (6.81)	23.92 (6.88)
SE1	35.55 (13.40)	20.58 (12.17)	29.51 (11.42)	13.71 (7.52)	29.54 (9.28)	21.93 (8.82)
BR2	19.74 (6.38)	29.42 (10.19)	23.62 (4.42)	26.38 (7.09)	21.67 (6.11)	23.87 (5.13)
SE2	23.64 (11.09)	26.70 (11.62)	25.06 (8.96)	20.95 (8.80)	20.72 (8.67)	20.80 (7.93)

Table B5
Experiment 4.

	Positive events			Negative events		
	Desirable trials	Undesirable trials	Overall	Desirable trials	Undesirable trials	Overall
Actual BR	30.66 (13.83)	23.46 (8.20)	27.06 (8.42)	24.37 (10.23)	38.78 (14.43)	31.58 (11.44)
BR1	25.47 (11.48)	30.04 (10.66)	27.75 (8.36)	31.00 (13.15)	31.94 (11.96)	31.47 (11.61)
SE1	28.23 (16.07)	35.73 (14.74)	31.98 (12.73)	25.64 (9.86)	26.94 (9.25)	26.29 (8.67)
SE2	28.50 (16.75)	31.33 (12.77)	29.92 (12.35)	24.25 (8.72)	30.42 (13.37)	27.34 (9.34)

Appendix C. Additional analyses – descriptive statistics by Event type and Desirability in Experiment 1

	Positive events		Negative events	
	Desirable trials	Undesirable trials	Desirable trials	Undesirable trials
Number of excluded capped trials	6.92 (3.28)	3.23 (1.92)	2.69 (2.93)	4.38 (3.93)
Memory error[a]	6.14 (3.47)	7.42 (2.98)	6.11 (2.04)	7.57 (2.66)
Vividness	3.68 (0.93)	3.76 (1.01)	3.34 (0.90)	3.41 (0.80)
Experience	1.94 (0.49)	2.10 (0.39)	1.31 (0.18)	1.41 (0.25)
Arousal	3.59 (0.71)	3.31 (0.86)	3.27 (1.19)	3.44 (1.25)
Magnitude of valence	3.94 (0.73)	4.03 (0.67)	4.31 (0.66)	4.50 (0.64)

Note. Standard deviations are shown in parentheses.
[a] Memory errors were calculated as the mean absolute difference between the recalled value and the actual probability.

Appendix D. Normative foundations and limitations of the revised update method, using the personal risk classification scheme

The basic logic of the update method, illustrated schematically in Fig. D1, runs as follows: For an event of given base rate, individuals in the population will over- or under-estimate that base rate to

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

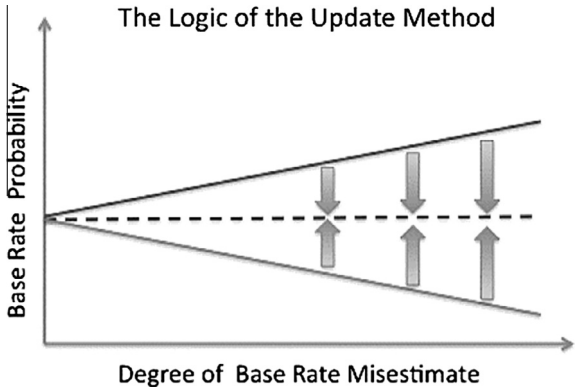

Fig. D1. For an event with true base rate of .4 (i.e., the event will ultimately be experienced by 40% of the population), indicated by the dashed line, individuals in the population will over- or under-estimate that base rate to a certain extent. On receipt of information about the true base rate, these individuals should adopt that new, correct base rate, thus changing their beliefs in the direction indicated by the arrows.

a certain extent. On receipt of information about the true base rate, these individuals should adopt that new, correct base rate, thus changing their beliefs in the direction indicated by the arrows. In the case of negative events, those over-estimating the base rate will receive 'desirable information' indicating that the true risk is lower, whereas those under-estimating will receive 'undesirable information'. For positive events, the valence of over- and under-estimates is reversed.

Of course, as outlined in the manuscript, participants will also frequently be in possession of individuating information, according to which they may conclude that, normatively, their individual risk lies above or below the (estimated) base rate: Someone not in possession of a bicycle is entitled to believe that their risk of experiencing a bicycle theft lies below the base rate, that is, the risk of the average person.

In a population of rational agents, who derive their own risk from the base rate in accordance with Bayes' theorem, the mean individual risk estimate will correspond to the base rate (see also Harris and Hahn, 2011), which is one of the reasons why Bayes' theorem is viewed as normative. Specifically, the mean individual risk estimate is the average across the responses of those people receiving a positive test result and those receiving a negative test result, for some diagnostic 'test' (individuating piece of knowledge such as family history) '*e*'. These averages are obtained by multiplying the respective posterior degrees of belief (Eqs. (D1) and (D2)) with the proportions of people expressing them (Eqs. (D3) and (D4)). It can be seen from Eqs. (D1) and (D3) and from Eqs. (D2) and (D4) that this multiplication process will cancel out the denominators in Eqs. (D1) and (D4), leaving the average response of the population equal to Eq. (D5). As $P(\neg e|h)$ and $P(e|h)$ must sum to 1, Eq. (7) reduces to $P(h)$, which equals the base rate.

$$P(h|e) = \frac{P(h)P(e|h)}{P(h)P(e|h) + P(\neg h)P(e|\neg h)} \tag{D1}$$

$$P(h|\neg e) = \frac{P(h)P(\neg e|h)}{P(\neg h)P(\neg e|\neg h) + P(h)P(\neg e|\neg h)} \tag{D2}$$

$$P(h)P(e|h) + P(\neg h)P(e|\neg h) \tag{D3}$$

$$P(h)P(\neg e|h) + P(\neg h)P(\neg e|\neg h) \tag{D4}$$

$$P(h)P(\neg e|h) + P(h)P(e|h) \tag{D5}$$

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

Because responses of individuals in positions of both positive and negative individuating knowledge average out in this way, the individuating knowledge of individuals cancels out.

However, a caveat arises from the fact that participants' base rate estimates are mistaken. As pointed out by Harris and Hahn (2011), individuating knowledge is dispensed 'by the world' and thus reflects the true, underlying base rate, not the base rate (erroneously) assumed by participants. This will give rise to systematic deviations between the mean base rate estimate and the mean estimate of individual risk. Where the base rate is over-estimated, the mean individual risk estimate will lie below the estimated base rate thus (falsely) suggesting individuals are 'optimistic'. Where the base rate is under-estimated, the mean individual risk estimate will lie above. These relationships are illustrated in the panel plots of Fig. D2.

Notably, it is a consequence of the bounded nature of the probability scale that the change scores for base rate over- and under-estimators who are in possession of diagnostic knowledge about their own individual risk only approximately balance out. Where that knowledge is very diagnostic, and the deviation from the true base rate sizeable, under-estimators who are otherwise entirely rational will exhibit larger amounts of absolute change in incorporating the new base rate (see bottom right hand panel of Fig. D2). For values of sensitivity and specificity that govern the diagnosticity of individuals' risk relevant knowledge or 'tests', where the likelihood ratio is 1, the update scores for over- and under-estimators will be exactly equal, as can be seen in the landscape plot Fig. D3 Panels A and B.

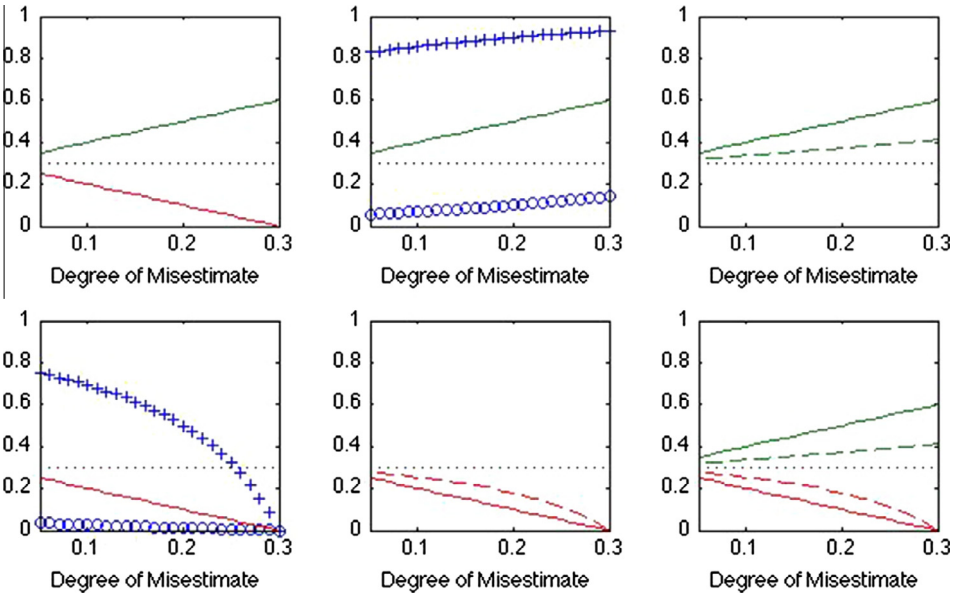

Fig. D2. The panel plots illustrate how groups of rational agents who over- and under-estimate the base rate will come to deviate from their respective population base rate (mis)estimates in their estimates of individual risk. The top left hand plot recapitulates Fig. D1, and represents the true base rate (dotted line), as well as over- and under-estimates of that base rate. Here the true base rate equals .3, over-and under-estimates range from .05 to .30 respectively. The middle panel of the top row shows for each level of base rate mis-estimate, the resulting estimates of individual risk for those in receipt of a positive test result ('+') and those in receipt of a negative test result ('o') for a hypothetical test of sensitivity, $P(e|h)$, = .7 and specificity, $P(\neg e|\neg h)$, = .7. Estimates of individual risk are derived using Eqs. (D1) and (D2). The top right panel shows the average (dashed line) of these estimates within the group (assuming representative sampling). As indicated in the text, this average lies below the group's mis-estimated base rate. Bottom left and middle panels show the corresponding plots for base rate under-estimators. The bottom right hand plot, finally, combines this information to generate the aggregate plot: as can be seen the mean individual risk estimates of over- and under-estimators (dashed lines) deviate from the over- and under-estimators assumed base rates.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

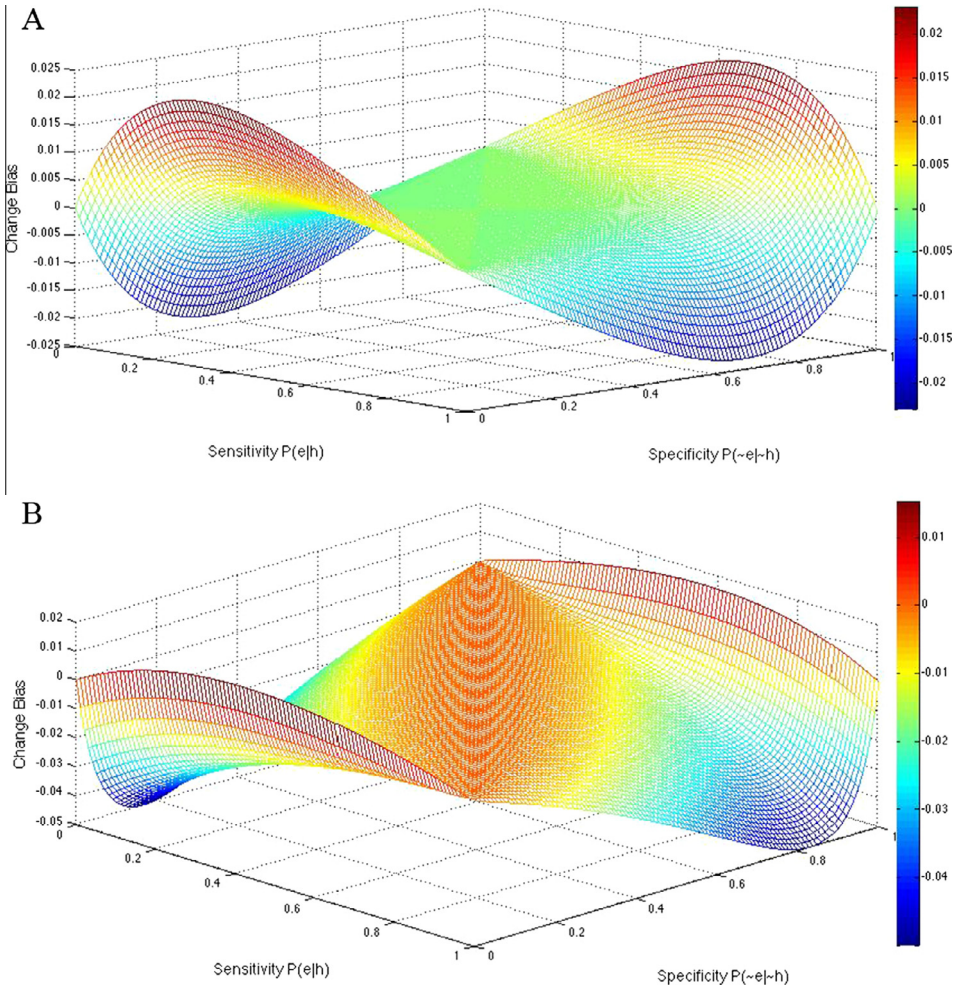

Fig. D3. The figure plots the extent to which the absolute update scores of rational agents who are mistaken only about the base rate differ between base rate over- and under-estimators. *x*- and *y*-axes represent sensitivity and specificity of the diagnostic 'test'. The *z* axis ('Change score bias') shows the degree to which the change scores of the under-estimators exceed those of the over-estimators. Panel A plots the space for a 'true' base rate = .50, under- and over-estimates deviate from this by ±.15. Panel B plots the corresponding space for a true base rate of .25 as in Table 1. Note also the smoothness in contrast with the space that arises with misclassifications as shown in Fig. 3.

Over a broad range of values, the deviation is very small, becoming sizeable only at very high levels of sensitivity and selectivity. Note also that those regions of the parameter space where the false positive rate exceeds the hit rate of the test (i.e., $P(\neg e|h) > P(e|h)$) imply a test whose meaning is reversed.

The inclusion of both negative and positive events guarantees that, even in the case where update scores are unequal in magnitude, the imbalance that arises for over- and under-estimates is un-confounded from what is psychologically 'desirable' and 'undesirable' information (see Fig. D4) – further highlighting the importance of including both positive and negative events in these studies.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

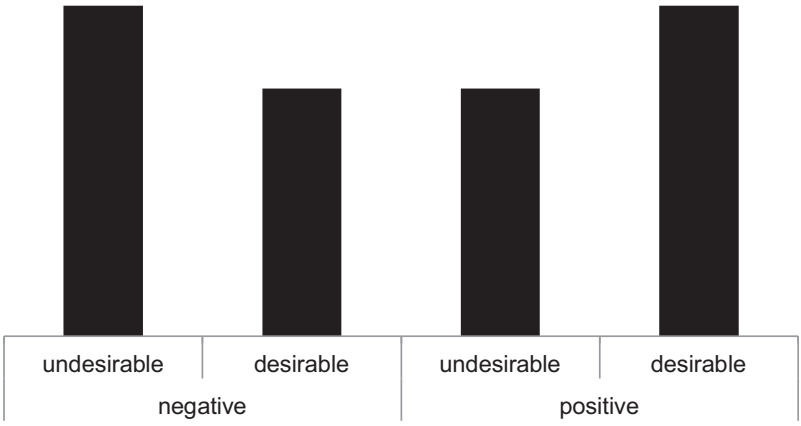

Fig. D4. Schematic illustration of outcomes of a change score test where absolute change scores deviate between rational over- and under-estimators of the base rate. In the case of negative events, under-estimates correspond to 'undesirable information'. However, for positive events, under-estimators will be receiving 'desirable information'.

Appendix E. Life events – Experiment 2

A healthy child born in the family
A warm sunny day in winter
Accidentally putting electronic device in washing machine
Attending a family member's wedding
Being applauded
Being arrested
Being asked on a date by someone you are attracted to
Being cheated on by partner
Being in unmanageable debt
Being offered a seat on a busy train
Being sent a gift from an anonymous sender
Being sworn at in public
Being told you are special
Being told you look attractive/handsome
Bone fracture
Buying a new home
Cancer (of digestive system/lung/prostate/breast/skin)
Card fraud
Chronic high blood pressure
Computer crash with permanent loss of all data
Constant healthy weight for 10 years
Death before 70
Death by infection
Dementia
Domestic burglary
Falling down stairs
Family throw you a surprise party
Finding £10 or more lying on the street
Finding something valuable you thought you had lost
Food poisoning with need to visit doctor

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

Fraud when buying something on the internet
Get trapped in a lift
Getting a large bonus payment at work
Getting a new pet animal
Getting an all-expenses paid holiday
Getting severely sunburnt
Getting unexpected tax break/benefit payment
Given complimentary dessert at a restaurant
Going to a movie premiere
Having a book/article published
Having a stroke
Hearing your favourite song in a public place (e.g. restaurant/shop)
Heart failure
Helped by a stranger when you need it
House vandalised
Living healthily past 80
Losing mobile phone
Losing wallet
Making a profit when selling a valuable item
Meet someone who likes the same things as you
Meeting a member of the royal family
Miss a flight
Miss the bus/train home
Mouse/rat in house
Not getting ill all winter
Parkinson's disease
Pet animal injured
Promotion without applying for one
Raise over £1000 for charity
Receiving an inheritance from unknown family member
Receiving an unexpected discount on a purchase
Seeing someone famous while on public transport
Severe injury due to accident (traffic or house)
Sexual dysfunction
Theft from person
Theft from vehicle
Unexpectedly meeting an old schoolmate
Upgraded to first class on a flight
Vehicle (car/bike) stolen
Victim of bullying at work (nonphysical)
Victim of mugging
Victim of violence by acquaintance
Victim of violence by stranger
Victim of violence with need to go to A&E
Winning a prize in a media (TV/radio/newspaper) competition
Winning a race
Winning a raffle at a fair
Winning an award in recognition of your work
Witnessing a traumatising accident
Your achievements in newspaper

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

E.1. Life events used during practice trials

Glaucoma
Winning the lottery

Appendix F. Life events and provided base rate statistics – Experiment 3

37 Negative events (provided risk for average person [percentage] in the right column):

Theft from a vehicle	76
Liver Disease	8
Dying before 80	41
Anxiety Disorder	13
Bone Fracture	38
Alzheimer's Disease	6
Heart Failure	31
Kidney Stones	10
Clinical obesity	30
Artificial Joint	11
Cancer of Digestive system	13
Chronic high blood pressure	20
Lung cancer	6
Type II diabetes	27
Dying before 70	18
Violence from a stranger	26
Severe insomnia	11
Spinal cord disease	13
Stroke	17
Prostate/breast cancer	5
Serious hearing problem	14
Violence from an acquaintance	28
Car theft	15
Abnormal heart rhythm	24
Dementia	13
Divorce	45
Gallbladder stones	10
Cancer	40
House vandalised	80
Hepatitis A/B	36
Tinnitus	8
Theft from person	45
Dying before 60	8
Alcoholism	7
Cataract	65
Skin cancer	3
Parkinson's Disease	4

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

19 Positive events:

Own own home	82
Marry a millionaire	3
Like first job after university	54
Have a starting salary greater than £40,000	7
Not spend a night in hospital in the next 5 years	55
Receive nationwide recognition within a profession	3
Have a starting salary greater than £20,000	55
Have an achievement recognised in the national press	3
Have a mentally gifted child	7
Visit the Amazonian rainforest	5
Be earning more than £80,000 in 10 years' time	2
Home's value doubles in any 5 year period	15
Have a starting salary greater than £30,000	16
Live past 90 years old	11
Receive a good job offer before graduating	17
Maintain a constant weight for the next 10 years	26
Last the whole of next winter without being ill	20
Graduate with a first	33
Have one's work recognised with an award	6

Appendix G. Life events and provided base rate statistics – Experiment 4

Positive	
Be earning more than £80,000 in 10 years' time	2
Marry a millionaire	3
Receive nationwide recognition within a profession	3
Have an achievement recognised in the national press	3
Visit the Amazonian rainforest	5
Have one's work recognised with an award	6
Have a starting salary greater than £40,000	7
Have a mentally gifted child	7
Live past 90 years old	11
Home's value doubles in any 5 year period	15
Have a starting salary greater than £30,000	16
Receive a good job offer before graduating	17
Last the whole of next winter without being ill	20
Graduate with a first	23
Maintain a constant weight for the next 10 years	26
Professional or managerial job after graduating	45
Like first job after university	54
Not spend a night in hospital in the next 5 years	55
Have a starting salary greater than £20,000	55
Own own home	82
Negative	
Skin cancer	3
Parkinson's Disease	4

(*continued on next page*)

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

Breast cancer	5
Lung cancer	6
Liver Disease	8
Dying before 60	8
Kidney Stones	10
Severe insomnia	11
Anxiety Disorder	13
Spinal cord disease	13
Dementia	13
Car theft	15
Chronic high blood pressure	20
Abnormal heart rhythm	24
Violence from a stranger	26
Type II diabetes	27
Clinical obesity	30
Cataract	65
Theft from a vehicle	76
House vandalised	80

References

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology, 68*, 804–825. http://dx.doi.org/10.1037//0022-3514.68.5.804.

Arnett, J. J. (2000). Optimistic bias in adolescent and adult smokers and nonsmokers. *Addictive Behaviors, 25*, 625–632. http://dx.doi.org/10.1016/S0306-4603(99)00072-6.

Babad, E., & Katz, Y. (1991). Wishful thinking – Against all odds. *Journal of Applied Social Psychology, 21*, 1921–1938. http://dx.doi.org/10.1111/j.1559-1816.1991.tb00514.x.

Bargh, J. A., & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 253–285). New York, NY: Cambridge University Press.

Bar-Hillel, M., & Budescu, D. (1995). The elusive wishful thinking effect. *Thinking & Reasoning, 1*, 71–103. http://dx.doi.org/10.1080/13546789508256906.

Bar-Hillel, M., Budescu, D. V., & Amar, M. (2008). Predicting World Cup results: Do goals seem more likely when they pay off? *Psychonomic Bulletin & Review, 15*, 278–283. http://dx.doi.org/10.3758/PBR.15.2.278.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck depression inventory-II*. San Antonio, TX: Psychological Corporation.

Benoit, J.-P., & Dubra, J. (2011). Apparent overconfidence. *Econometrica, 79*, 1591–1625. http://dx.doi.org/10.3982/ECTA8583.

Benoit, J.-P., Dubra, J., & Moore, D. A. (2015). Does the better-than-average effect show that people are overconfident? Two experiments. *Journal of the European Economic Association, 13*, 293–329. http://dx.doi.org/10.1111/jeea.12116.

Bilgin, B. (2012). Losses loom more likely than gains: Propensity to imagine losses increases their subjective probability. *Organizational Behavior and Human Decision Processes, 118*, 203–215. http://dx.doi.org/10.1016/j.obhdp.2012.03.008.

Bleumink, G. S., Knetsch, A. M., Sturkenboom, M. C., Straus, S. M., Hofman, A., Deckers, J. W., ... Stricker, B. H. C. (2004). Quantifying the heart failure epidemic: Prevalence, incidence rate, lifetime risk and prognosis of heart failure. *European Heart Journal, 25*, 1614–1619. http://dx.doi.org/10.1016/j.ehj.2004.06.038.

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: The role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological Bulletin, 130*, 813–838. http://dx.doi.org/10.1037/0033-2909.130.5.813.

Chambers, J. R., Windschitl, P. D., & Suls, J. (2003). Egocentrism, event frequency, and comparative optimism: When what happens frequently is "more likely to happen to me". *Personality & Social Psychology Bulletin, 29*, 1343–1356. http://dx.doi.org/10.1177/0146167203256870.

Chowdhury, R., Sharot, T., Wolfe, T., Düzel, E., & Dolan, R. J. (2014). Optimistic update bias increases in older age. *Psychological Medicine, 44*(09), 2003–2012. http://dx.doi.org/10.1017/S0033291713002602.

Christensen-Szalanski, J. J. J., Beck, D. E., Christensen-Szalanski, C. M., & Koepsell, T. D. (1983). Effects of expertise and experience on risk judgments. *Journal of Applied Psychology, 68*, 278–284. http://dx.doi.org/10.1037/0021-9010.68.2.278.

Colvin, C. R., & Block, J. (1994). Do positive illusions foster mental health? An examination of the Taylor and Brown formulation. *Psychological Bulletin, 116*, 3–20. http://dx.doi.org/10.1037/0033-2909.116.1.3.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58*, 1–73. http://dx.doi.org/10.1016/0010-0277(95)00664-8.

Critcher, C. R., & Dunning, D. (2009). How chronic self-views influence (and mislead) self-assessments of task performance: Self-views shape bottom-up experiences with the task. *Journal of Personality and Social Psychology, 97*, 931–945. http://dx.doi.org/10.1037/a0017452.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

de Molière, L., & Harris, A. J. L. (2016). Conceptual and direct replications fail to support the Stake-Likelihood Hypothesis as an explanation for the interdependence of utility and likelihood judgments. *Journal of Experimental Psychology: General, 145*, e13–e26. http://dx.doi.org/10.1037/xge0000124.

Ditto, P. H., Jemmott, J. B., & Darley, J. M. (1988). Appraising the threat of illness: A mental representational approach. *Health Psychology, 7*, 183–201. http://dx.doi.org/10.1037/0278-6133.7.2.183.

Ditto, P., & Lopez, D. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology, 63*, 568–584. http://dx.doi.org/10.1037/0022-3514.63.4.568.

Dunning, D., Heath, C., & Suls, J. (2004). Flawed self-assessment. *Psychological Science in the Public Interest, 5*, 71–106. http://dx.doi.org/10.1111/j.1529-1006.2004.00018.x.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57*, 1082–1090. http://dx.doi.org/10.1037/0022-3514.57.6.1082.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York, NY: Wiley.

Eil, D., & Rao, J. M. (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics, 3*, 114–138. http://dx.doi.org/10.1257/mic.3.2.114.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review, 101*, 519–527. http://dx.doi.org/10.1037/0033-295X.101.3.519.

Galesic, M., Olsson, H., & Rieskamp, J. (2012). Social sampling explains apparent biases in judgments of social environments. *Psychological Science, 23*, 1515–1523. http://dx.doi.org/10.1177/0956797612445313.

Garrett, N., & Sharot, T. (2014). How robust is the optimistic update bias for estimating self-risk and population base rates. *PLoS ONE, 9*(6), e98848. http://dx.doi.org/10.1371/journal.pone.0098848.

Garrett, N., Sharot, T., Faulkner, P., Korn, C. W., Roiser, J. P., & Dolan, R. J. (2014). Losing the rose tinted glasses: Neural substrates of unbiased belief updating in depression. *Frontiers in Human Neuroscience, 8*. http://dx.doi.org/10.3389/fnhum.2014.00639.

Gibson, B., & Sanbonmatsu, D. M. (2004). Optimism, pessimism, and gambling: The downside of optimism. *Personality and Social Psychology Bulletin, 30*, 149–160. http://dx.doi.org/10.1177/0146167203259929.

Hahn, U., & Harris, A. J. L. (2014). What does it mean to be biased: Motivated reasoning and rationality. *The Psychology of Learning and Motivation*. http://dx.doi.org/10.1016/B978-0-12-800283-4.00002-2.

Hardman, D. (2009). *Judgment and decision making: Psychological perspectives*. Chichester, UK: BPS Blackwell.

Harris, A. J. L. (2009). *Investigating the influence of outcome utility on estimates of probability* (Unpublished doctoral dissertation). Cardiff, Wales: Cardiff University.

Harris, A. J. L., Corner, A., & Hahn, U. (2009). Estimating the probability of negative events. *Cognition, 110*, 51–64. http://dx.doi.org/10.1016/j.cognition.2008.10.006.

Harris, D. M., & Guten, S. (1979). Health-protective behavior: An exploratory study. *Journal of Health and Social Behavior, 20*, 17–29. http://dx.doi.org/10.2307/2136475.

Harris, A. J. L., & Hahn, U. (2011). Unrealistic optimism about future life events: A cautionary note. *Psychological Review, 118*, 135–154. http://dx.doi.org/10.1037/a0020997.

Harris, A. J. L., Shah, P., Catmur, C., Bird, G., & Hahn, U. (2013). Autism, optimism and positive events: Evidence against a general optimistic bias. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Hawe, E., Talmud, P. J., Miller, G. J., & Humphries, S. E. (2003). Family history is a coronary heart disease risk factor in the Second Northwick Park Heart Study. *Annals of Human Genetics, 67*, 97–106. http://dx.doi.org/10.1046/j.1469-1809.2003.00017.x.

Helweg-Larsen, M., & Shepperd, J. A. (2001). Do moderators of the optimistic bias affect personal or target risk estimates? A review of the literature. *Personality and Social Psychology Review, 5*, 74–95. http://dx.doi.org/10.1207/S15327957PSPR0501_5.

Hertwig, R., Pachur, T., & Kurzenhäuser, S. (2005). Judgments of risk frequencies: Tests of possible cognitive mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 621–642. http://dx.doi.org/10.1037/0278-7393.31.4.621.

HM Treasury (n.d.). *The Green Book: Appraisal and evaluation in central government*. <http://www.hm-treasury.gov.uk/data_greenbook_index.htm> Retrieved September 16, 2015.

Hogarth, R. M., & Karelaia, N. (2012). Entrepreneurial success and failure: Confidence and fallible judgment. *Organization Science, 23*, 1733–1747. http://dx.doi.org/10.1287/orsc.1110.0702.

Irwin, F. W. (1953). Stated expectations as functions of probability and desirability of outcomes. *Journal of Personality, 21*, 329–335. http://dx.doi.org/10.1111/j.1467-6494.1953.tb01775.x.

Juslin, P., Nilsson, H., & Winman, A. (2009). Probability theory, not the very guide of life. *Psychological Review, 116*, 856–874. http://dx.doi.org/10.1037/a0016979.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251. http://dx.doi.org/10.1037/h0034747.

Korn, C. W., Prehn, K., Park, S. Q., Walter, H., & Heekeren, H. R. (2012). Positively biased processing of self-relevant social feedback. *Journal of Neuroscience, 21*, 16832–16844. http://dx.doi.org/10.1523/JNEUROSCI.3016-12.2012.

Korn, C. W., Sharot, T., Walter, H., Heekeren, H. R., & Dolan, R. J. (2014). Depression is related to an absence of optimistically biased belief updating about future life events. *Psychological Medicine, 44*, 579–592. http://dx.doi.org/10.1017/S0033291713001074.

Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological Bulletin, 133*, 95–121. http://dx.doi.org/10.1037/0033-2909.133.1.95.

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77*, 221–232. http://dx.doi.org/10.1037/0022-3514.77.2.221.

Kruger, J., & Burrus, J. (2004). Egocentrism and focalism in unrealistic optimism (and pessimism). *Journal of Experimental Social Psychology, 40*, 332–340. http://dx.doi.org/10.1016/j.jesp.2003.06.002.

Kruger, J., Windschitl, P. D., Burrus, J., Fessel, F., & Chambers, J. R. (2008). The rational side of egocentrism in social comparisons. *Journal of Experimental Social Psychology, 44*, 220–232. http://dx.doi.org/10.1016/j.jesp.2007.04.001.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480–498.
Kuzmanovic, B., Jefferson, A., & Vogeley, K. (2015). Self-specific optimism bias in belief updating is associated with high trait optimism. *Journal of Behavioral Decision Making, 28*, 281–293. http://dx.doi.org/10.1002/bdm.1849.
Kuzmanovic, B., Jefferson, A., & Vogeley, K. (2016). The role of the neural reward circuitry in self-referential optimistic belief updates. *NeuroImage, 133*, 151–162. http://dx.doi.org/10.1016/j.neuroimage.2016.02.014.
Lench, H. C. (2009). Automatic optimism: The affective basis of judgments about the likelihood of future events. *Journal of Experimental Psychology: General, 138*, 187–200. http://dx.doi.org/10.1037/a0015380.
Lench, H. C., & Bench, S. W. (2012). Automatic optimism: Why people assume their futures will be bright. *Social and Personality Psychology Compass, 6*, 347–360. http://dx.doi.org/10.1111/j.1751-9004.2012.00430.x.
Lench, H. C., & Ditto, P. H. (2008). Automatic optimism: Biased use of base rate information for positive and negative events. *Journal of Experimental Social Psychology, 44*, 631–639. http://dx.doi.org/10.1016/j.jesp.2007.02.011.
Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge, UK: Cambridge University Press.
Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 551–578. http://dx.doi.org/10.1037/0278-7393.4.6.551.
Mannes, A. E., & Moore, D. A. (2013). A behavioral demonstration of overconfidence in judgment. *Psychological Science, 24*, 1190–1197.
Massey, C., Simmons, J. P., & Armor, D. A. (2011). Hope over experience: Desirability and the persistence of optimism. *Psychological Science, 22*, 274–281. http://dx.doi.org/10.1177/0956797610396223.
Merkle, C., & Weber, M. (2011). True overconfidence: The inability of rational information processing to account for apparent overconfidence. *Organizational Behavior and Human Decision Processes, 116*, 262–271. http://dx.doi.org/10.1016/j.obhdp.2011.07.004.
Miles, S., & Scaife, V. (2003). Optimistic bias and food. *Nutrition Research Reviews, 16*, 3–19. http://dx.doi.org/10.1079/NRR200249.
Mobius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2011). *Managing self-confidence: Theory and experimental evidence*. Working paper.
Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*, 502–517. http://dx.doi.org/10.1037/0033-295X.115.2.502.
Moore, D. A., & Small, D. A. (2007). Error and bias in comparative judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology, 92*, 972–989. http://dx.doi.org/10.1037/0022-3514.92.6.972.
Moore, D. A., & Small, D. A. (2008). When it is rational for the majority to believe that they are better than average. In J. I. Krueger (Ed.), *Rationality and social responsibility: Essays in honor of Robyn Mason Dawes* (pp. 141–174). New York, NY: Psychology Press.
Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in Cognitive Sciences, 14*, 435–440. http://dx.doi.org/10.1016/j.tics.2010.07.004.
Moutsiana, C., Garrett, N., Clarke, R. C., Beau Lotto, R., Blakemore, S.-J., & Sharot, T. (2013). Human development of the ability to learn from bad news. *Proceedings of the National Academy of Sciences, 110*, 16396–16401. http://dx.doi.org/10.1073/pnas.1305631110.
Oaksford, M. (2015). Imaging deductive reasoning and the new paradigm. *Frontiers in Human Neuroscience, 9*. http://dx.doi.org/10.3389/fnhum.2015.00101 101.
Office for National Statistics (2000). Registrations of cancer diagnosed in 1994–1997, England and Wales. *Health Statistics Quarterly, 7*, 71–82.
Pfeifer, P. E. (1994). Are we overconfident in the belief that probability forecasters are overconfident? *Organizational Behavior and Human Decision Processes, 58*, 203–213. http://dx.doi.org/10.1006/obhd.1994.1034.
Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology, 72*, 346–354.
Pruitt, D. G., & Hoge, R. D. (1965). Strength of the relationship between the value of an event and its subjective probability as a function of the method of measurement. *Journal of Experimental Psychology, 69*, 483–489. http://dx.doi.org/10.1037/h0021721.
Puri, M., & Robinson, D. T. (2007). Optimism and economic choice. *Journal of Financial Economics, 86*, 71–99. http://dx.doi.org/10.1016/j.jfineco.2006.09.003.
Risen, J. L., & Gilovich, T. (2007). Another look at why people are reluctant to exchange lottery tickets. *Journal of Personality and Social Psychology, 93*, 12–22. http://dx.doi.org/10.1037/0022-3514.93.1.12.
Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review, 117*, 1144–1167. http://dx.doi.org/10.1037/a0020511.
Shah, P. (2012). Toward a neurobiology of unrealistic optimism. *Frontiers in Psychology, 3*. http://dx.doi.org/10.3389/fpsyg.2012.00344.
Sharot, T. (2012). *The Optimism Bias: Why we're wired to look on the bright side*. London, UK: Constable & Robinson Limited.
Sharot, T., Guitart-Masip, M., Korn, C. W., Chowdhury, R., & Dolan, R. J. (2012). How dopamine enhances an optimism bias in humans. *Current Biology, 22*, 1477–1481. http://dx.doi.org/10.1016/j.cub.2012.05.053.
Sharot, T., Kanai, R., Marston, D., Korn, C. W., Rees, G., & Dolan, R. J. (2012). Selectively altering belief formation in the human brain. *Proceedings of the National Academy of Sciences of the United States of America, 109*, 17058–17062. http://dx.doi.org/10.1073/pnas.1205828109.
Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience, 14*, 1475–1479. http://dx.doi.org/10.1038/nn.2949.
Sharot, T., Riccardi, A. M., Raio, C. M., & Phelps, E. A. (2007). Neural mechanisms mediating optimism bias. *Nature, 450*, 102–105. http://dx.doi.org/10.1038/nature06280.
Shepherd, R. (2002). Resistance to changes in diet. *Proceedings of the Nutrition Society, 61*, 267–272. http://dx.doi.org/10.1079/PNS2002147.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004

Shepperd, J. A., Klein, W. M., Waters, E. A., & Weinstein, N. D. (2013). Taking stock of unrealistic optimism. *Perspectives on Psychological Science, 8*, 395–411. http://dx.doi.org/10.1177/1745691613485247.

Simmons, J. P., & Massey, C. (2012). Is optimism real? *Journal of Experimental Psychology: General, 141*, 630–634. http://dx.doi.org/10.1037/a0027405.

Soll, J. B. (1996). Determinants of overconfidence and miscalibration: The roles of random error and ecological structure. *Organizational Behavior and Human Decision Processes, 65*, 117–137. http://dx.doi.org/10.1006/obhd.1996.0011.

Strunk, D. R., Lopez, H., & DeRubeis, R. J. (2006). Depressive symptoms are associated with unrealistic negative predictions of future life events. *Behaviour Research and Therapy, 44*, 861–882. http://dx.doi.org/10.1016/j.brat.2005.07.001.

Sunstein, C. R. (2000). *Behavioral law and economics*. Cambridge, UK: Cambridge University Press.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica, 47*, 143–148. http://dx.doi.org/10.1016/0001-6918(81)90005-6.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*, 193–210. http://dx.doi.org/10.1037/0033-2909.103.2.193.

Taylor, S. E., & Brown, J. D. (1994). Positive illusions and well-being revisited: Separating fact from fiction. *Psychological Bulletin, 116*(1), 21–27.

Tenney, E. R., Logg, J. M., & Moore, D. A. (2015). (Too) optimistic about optimism: The belief that optimism improves performance. *Journal of Personality and Social Psychology, 108*, 377–399.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12*, 435–502.

Thagard, P. (2000). *Coherence in thought and action*. Cambridge, MA: MIT Press.

van den Steen, E. (2004). Rational overoptimism (and other biases). *American Economic Review, 94*, 1141–1151. http://dx.doi.org/10.1257/0002828042002697.

van der Velde, F. W., Hooykaas, C., & van der Joop, P. (1992). Risk perception and behavior: Pessimism, realism, and optimism about aids-related health behavior. *Psychology & Health, 6*, 23–38. http://dx.doi.org/10.1080/08870449208402018.

van der Velde, F. W., van der Pligt, J., & Hooykaas, C. (1994). Perceiving AIDS-related risk: Accuracy as a function of differences in actual risk. *Health Psychology, 13*, 25–33. http://dx.doi.org/10.1037/0278-6133.13.1.25.

Vosgerau, J. (2010). How prevalent is wishful thinking? Misattribution of arousal causes optimism and pessimism in subjective probabilities. *Journal of Experimental Psychology: General, 139*, 32–48. http://dx.doi.org/10.1037/a0018144.

Weber, E. U. (1994). From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin, 115*, 228–242. http://dx.doi.org/10.1037/0033-2909.115.2.228.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39*, 806–820. http://dx.doi.org/10.1037/0022-3514.39.5.806.

Weinstein, N. D. (1982). Unrealistic optimism about susceptibility to health problems. *Journal of Behavioral Medicine, 5*, 441–460. http://dx.doi.org/10.1007/BF00845372.

Weinstein, N. D. (1984). Why it won't happen to me: Perceptions of risk factors and susceptibility. *Health Psychology, 3*, 431–457. http://dx.doi.org/10.1037/0278-6133.3.5.431.

Weinstein, N. D. (1987). Unrealistic optimism about susceptibility to health problems: Conclusions from a community-wide sample. *Journal of Behavioral Medicine, 10*, 481–500. http://dx.doi.org/10.1007/BF00846146.

Weinstein, N. D. (1989). Effects of personal experience on self-protective behavior. *Psychological Bulletin, 105*, 31–50. http://dx.doi.org/10.1037/0033-2909.105.1.31.

Weinstein, N. D. (1998). Accuracy of smokers' risk perceptions. *Annals of Behavioral Medicine, 20*, 135–140. http://dx.doi.org/10.1007/BF02884459.

Weinstein, N. D., & Klein, W. M. (1996). Unrealistic optimism: Present and future. *Journal of Social and Clinical Psychology, 15*, 1–8. http://dx.doi.org/10.1521/jscp.1996.15.1.1.

Welsh, M. B., & Navarro, D. J. (2012). Seeing is believing: Priors, trust, and base rate neglect. *Organizational Behavior and Human Decision Processes, 119*, 1–14.

Windschitl, P. D., Smith, A. R., Rose, J. P., & Krizan, Z. (2010). The desirability bias in predictions: Going optimistic without leaving realism. *Organizational Behavior and Human Decision Processes, 111*, 33–47. http://dx.doi.org/10.1016/j.obhdp.2009.08.003.

Wiswall, M., & Zafar, B. (2015). How do college students respond to public information about earnings? *Journal of Human Capital, 9*(2), 117–169. http://dx.doi.org/10.1086/681542.

Please cite this article in press as: Shah, P., et al. A pessimistic view of optimistic belief updating. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.05.004